

Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
BSc (Hons) Aaron Tan Wai Kit
born in Kuala Lumpur, Malaysia
Oral examination: 13 December 2019

Structural insights into HIV-1 capsid assembly, maturation and stability by cryo-electron tomography

Referees: Dr. Martin Beck

Prof. Dr. Britta Brügger

I dedicate this thesis to my parents, Moses and Jolyn.

Abstract

Human immunodeficiency virus type 1 (HIV-1) is an enveloped lentivirus from the family Retroviridae which infects CD4⁺ T-lymphocytes in a human host, leading to Acquired Immunodeficiency Syndrome (AIDS) if untreated. A subset of retroviruses, most notably lentiviruses such as HIV-1, are unique in their ability to infect non-dividing cells. To do this, the reverse transcribed viral genome must be trafficked across an intact nuclear membrane and integrated into the host cell genome. The viral capsid plays a central role in this process.

The first stage of capsid assembly is polymerisation of the viral polyprotein Gag via its CA (capsid) domain into a hexagonal immature lattice, forming a truncated sphere. The viral protease cleaves Gag and frees the CA domain, which rearranges to form a conical capsid around the viral genome, built from CA hexamers and pentamers. Despite advances in recent years, many open questions remain about immature Gag lattice assembly, maturation and modulation of capsid stability by host factors upon infection.

One key question that has persisted in the field is how the remarkable structural transition between the immature Gag lattice and the mature CA lattice is achieved, which involves breaking almost all of the interactions stabilising the immature lattice. To address this, I applied cryo-electron tomography (cryo-ET) and subtomogram averaging to obtain high resolution structures of immature and mature CA in a panel of HIV-1 constructs containing different combinations of proteolytic cleavage sites inactivated by mutation. Unexpectedly, proteolytic processing directly on either side of CA was sufficient for mature lattice formation at low frequencies. I also show that a beta-hairpin domain at the CA N-terminus, previously proposed to be a structural switch, is dispensable for maturation. Instead, destabilisation of a six-helix bundle between the CA C-terminus and the adjacent SP1 peptide is the main structural determinant of maturation.

Viral maturation is tightly linked to immature Gag lattice assembly, but many details such as the basic unit of lattice assembly remain unclear. The immature lattice is maintained by inter- and intra-hexamer interactions but is not a complete sphere, and the structure of Gag at discontinuous lattice edges is unknown. I implemented a new workflow to obtain Gag lattice structures by subtomogram classification of a cryo-ET data set of intact HIV-1 virions. These structures show that Gag forms novel, incomplete hexamers at lattice edges and that the CA-SP1 region forms ordered helical bundles in partial hexamers. Molecular dynamics simulations suggest that these partial bundles exhibit increased an tendency to unfold, suggesting a role of partial hexamer structures in initiation of maturation.

Capsid stability after cell entry is important to prevent degradation of the viral RNA genome, and is modulated by small molecules such as inositol hexakisphosphate (IP6) and host proteins, including cleavage and polyadenylation specific factor 6 (CPSF6) and nucleoporin 153 (Nup153). A combination of CA pentamers and hexamers that flex to adopt different curvatures provides many different potential interfaces for cofactor binding. I developed a workflow to routinely obtain near-atomic resolution structures of CA hexamers and pentamers, by subtomogram averaging of conical, IP6-stabilised *in vitro* CA assemblies. These were used to investigate CPSF6 and Nup153 binding to pentamers and the effect of lattice curvature on the common binding pocket for these factors. The structures obtained show that CPSF6 and Nup153 do not bind to pentamers at the concentrations used, and that lattice flexibility can modulate Nup153 binding to hexamers.

Zusammenfassung

Das humane Immunodefizienz-Virus Type 1 (HIV-1) ist ein Membran-umhülltes Lentivirus und gehört zu der Familie der Retroviridae. HIV-1 infiziert CD4⁺ T-Lymphozyten, was zum Ausbruch des erworbenen Immunschwäche-Syndroms (AIDS) führen kann. Einige Retroviren, besonders die Lentiviren zum Beispiel HIV-1, besitzen die einzigartige Fähigkeit nicht-teilende Zellen zu infizieren. Dazu muss das revers-transkribierte virale Genom durch die intakte Zellkernmembran transportiert und in das Wirtszellgenom integriert werden. Das virale Kapsid spielt eine wichtige Rolle in diesem Prozess.

In der ersten Phase des Kapsidaufbaus polymerisiert das virale Polyprotein Gag über seine capsid (CA) Domäne in ein hexagonales Gitter, welches eine unvollständige Kugel unterhalb der Virushülle formt. Zur Reifung der Kapsidstruktur schneidet die virale Protease das Protein Gag, ein Prozess, durch den die CA Domäne frei wird. Die CA Domäne reorganisiert sich daraufhin und bildet dann eine konische Struktur aus CA Hexameren und Pentameren um das Genom herum. Trotz des Fortschritts in den letzten Jahren, bleiben viele Fragen bezüglich der initialen Kapsidassemblierung, der Reifung und Modulierung der Kapsidstabilität durch Wirtsfaktoren während der Infektion unbeantwortet.

Eine Kernfrage die nach wie vor diskutiert wird, ist, wie die dramatische, strukturelle Umorganisation, von der unreifen, hexagonale Gag Gitter Struktur zu der reifen konischen CA Struktur abläuft, da beinahe alle Interaktionen der unreifen Gag Struktur unterbrochen werden. Um diese Frage zu beantworten, habe ich Kryoelektronentomographie (KryoET) und "Subtomogram Averaging" verwendet, um hochaufgelöste Strukturen von unreifen und reifen CA-Strukturen in einem Set an HIV-1 Mutanten, bei denen verschiedene proteolytische Schnittstellen in dem Polyprotein Gag maskiert wurden, zu bestimmen. Unerwarteter Weise hat sich gezeigt, dass ein proteolytischer Schnitt, an je einer Seite von CA ausreichend ist, um die Kapsidreifung zu induzieren. Außerdem zeige ich hier, dass eine Beta-Haarnadel Domäne am CA N-Terminus, welcher davor als Schalter der strukturellen Umorganisation vermutet wurde, für die Reifung verzichtbar ist. Stattdessen scheint die Destabilisierung des sechs Helix Bündels zwischen der CA C-terminus und des daneben liegenden SP1 Peptids der wichtigste Faktor für die strukturelle Veränderung zu sein.

Virusreifung ist eng mit der Assemblierung des Gag Gitters verbunden. Viele Einzelheiten, wie die Grundeinheit der unreifen, Gag Gitter Struktur sind unbekannt. Das unreife, hexagonale Gag Gitter wird durch die Interaktionen in und zwischen den Gag Hexameren stabilisiert. Das Gag Gitter formt allerdings keine vollständige Kugel und die Gag Struktur an den Gitterkanten ist bisher nicht bekannt. Ich habe daher eine neue Methodik etabliert, um Gag Gitter Strukturen mit Hilfe von Subtomogram Klassifizierung aus KryoET Daten in intakten HIV-1 Partikeln zu bestimmen. Die erhaltenden Strukturen zeigen, dass Gag bisher unbekannte, unvollständige Hexamere an Gitterkanten bildet und dass die CA-SP1 Region geordnete Helix Bündel in partiellen Hexameren ausbildet. Molecular Dynamics Simulationen lassen vermuten, dass diese partielle Helix Bündel eine erhöhte Wahrscheinlichkeit vorweisen sich zu entfalten und damit eine Rolle bei der Initiation der Reifung einnehmen könnten.

Kapsid Stabilität nach Zelleintritt ist wichtig, um den Abbau des viralen RNA Genomes zu verhindern. Die Kapsidstabilität wiederum wird von kleinen Molekülen wie Inositol hexakisphosphat und Wirtsfaktoren wie cleavage and polyadenylation specific factor 6 (CPSF6) und nucleoporin 153 (Nup153) moduliert. Eine Kombination von flexiblen CA Pentameren und Hexameren in der reifen, konischen CA Struktur, die es der Struktur erlaubt verschiedene Krümmungen anzunehmen, bietet viele mögliche Bindestellen für Ko-Faktoren. Ich habe eine Methode entwickelt, um Routine-mäßig hochaufgelöste Strukturen mit fast atomarer Auflösung von CA Hexameren und CA Pentameren durch Subtomogram Averaging direkt aus konischen, IP6-stabilisierten CA Kapsids, die in-vitro assembliert wurden, zu erhalten. Aus diesen Strukturen können wir nun verstehen wie CPSF6 und Nup153 an Pentamere bindet und welche Effekte diese Interaktion auf die CA Krümmung und die Ko-Faktor Bindestellen haben. Die erhaltenden Strukturen zeigen, dass CPSF6 und Nup153, in den verwendeten Konzentrationen nicht an CA Pentamere bindet und dass die Gitter Flexibilität Nup153 Interaktion an CA Hexameren modulieren kann.

The work presented in this thesis was carried out in the laboratory of Dr. John Briggs, at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany and at the MRC Laboratory of Molecular Biology in Cambridge, United Kingdom.

Contents

| | |
|--|-------------|
| Abstract | i |
| Zusammenfassung | iii |
| Contents | vii |
| List of Figures | xiii |
| List of Tables | xvii |
| Abbreviations | xix |
| 1 Introduction | 1 |
| 1.1 Human immunodeficiency virus type 1 (HIV-1) | 1 |
| 1.1.1 Organisation of the HIV-1 genome | 2 |
| 1.1.2 HIV-1 proteins | 3 |
| 1.1.3 HIV-1 infection and replication | 10 |
| 1.1.4 HIV-1 assembly | 11 |
| 1.1.5 HIV-1 maturation | 12 |
| 1.1.6 HIV-1 fusion and nuclear entry | 14 |
| 1.2 Cryo-electron tomography (cryo-ET) and subtomogram averaging | 15 |

vii

| | | |
|----------|--|-----------|
| 1.2.1 | Electron microscopy of vitrified biological samples | 15 |
| 1.2.2 | Tomography and three-dimensional reconstruction from projections | 19 |
| 1.2.3 | Subtomogram averaging | 21 |
| 1.2.4 | Image classification in cryo-EM | 23 |
| 2 | The structural mechanism of HIV-1 maturation | 27 |
| 2.1 | Introduction | 28 |
| 2.1.1 | HIV-1 maturation | 28 |
| 2.2 | Aims | 31 |
| 2.3 | Materials and Methods | 31 |
| 2.3.1 | Production and purification of mutant virus particles | 31 |
| 2.3.2 | Cryo-EM grid preparation | 33 |
| 2.3.3 | Cryo-electron tomography | 33 |
| 2.3.4 | Image processing | 35 |
| 2.3.5 | Visualisation of structures and maps | 38 |
| 2.4 | Results | 40 |
| 2.4.1 | Large data sets reveal novel minority phenotypes | 40 |
| 2.4.2 | Morphological analysis of HIV-1 Gag cleavage mutants | 40 |
| 2.4.3 | Structures of immature CA in cleavage mutants | 44 |
| 2.4.4 | Structures of mature CA in cleavage mutants | 46 |
| 2.5 | Discussion | 49 |
| 2.5.1 | The complex relationship between proteolytic and structural maturation | 49 |
| 2.5.2 | The beta-hairpin is dispensable for maturation | 50 |
| 2.5.3 | CA-SP1 bundle unfolding is the structural switch for maturation . | 51 |

| | | |
|----------|--|-----------|
| 2.5.4 | A model for HIV-1 maturation | 51 |
| 3 | Immature HIV-1 Gag assembly and initiation of CA maturation | 53 |
| 3.1 | Introduction | 54 |
| 3.1.1 | Immature HIV-1 Gag lattice assembly | 54 |
| 3.2 | Aims | 55 |
| 3.3 | Materials and Methods | 56 |
| 3.3.1 | Cryo-ET data set of purified immature HIV-1 viruses | 56 |
| 3.3.2 | Removal of duplicate and misaligned subtomogram positions | 56 |
| 3.3.3 | Initial geometric selection of lattice edges | 57 |
| 3.3.4 | Further alignment prior to classification | 58 |
| 3.3.5 | Image-based subtomogram classification approaches | 59 |
| 3.3.6 | Molecular dynamics simulations | 62 |
| 3.4 | Results | 63 |
| 3.4.1 | Classification of Gag hexamer species at the lattice edge | 63 |
| 3.4.2 | Dimer and trimer interfaces in incomplete Gag hexamers | 65 |
| 3.4.3 | CA-SP1 junction integrity in incomplete hexamers | 67 |
| 3.5 | Discussion | 69 |
| 3.5.1 | Gag dimers are likely to be the basic immature lattice assembly unit | 69 |
| 3.5.2 | The CA-SP1 region still forms partial helical bundles in incomplete hexamers | 72 |
| 3.5.3 | Stability of the immature lattice in the context of retroviral evolution | 73 |
| 4 | Modulation of HIV-1 capsid stability and nuclear import by host cofactors | 75 |
| 4.1 | Introduction | 76 |

| | | |
|----------|---|------------|
| 4.1.1 | Inositol hexakisphosphate (IP ₆) | 77 |
| 4.1.2 | Nuclear import cofactors | 77 |
| 4.1.3 | <i>In vitro</i> conical assemblies of HIV-1 CA with IP ₆ | 79 |
| 4.2 | Aims | 79 |
| 4.3 | Materials and Methods | 80 |
| 4.3.1 | Expression, purification and <i>in vitro</i> assembly of conical CA cores with IP ₆ | 80 |
| 4.3.2 | Negative staining of VLPs for morphology experiments | 81 |
| 4.3.3 | Cryo-EM grid preparation | 81 |
| 4.3.4 | Cryo-electron tomography of VLPs | 82 |
| 4.3.5 | Image processing | 83 |
| 4.4 | Results | 89 |
| 4.4.1 | <i>In vitro</i> CA assemblies with IP ₆ are identical in structure to viral capsids | 89 |
| 4.4.2 | Presence of IP ₆ in the hexamer and pentamer structures | 93 |
| 4.4.3 | CPSF6 binds to CA hexamers but not pentamers | 99 |
| 4.4.4 | Nup153 also binds to hexamers but not pentamers | 101 |
| 4.5 | Discussion | 106 |
| 4.5.1 | Modulation of capsid stability | 106 |
| 4.5.2 | Nuclear import cofactors Nup153 and CPSF6 | 107 |
| 4.5.3 | Implications for therapeutic design | 109 |
| 5 | Conclusions and perspectives | 111 |
| 5.1 | Structural switch controlling HIV-1 maturation | 112 |
| 5.2 | HIV-1 lattice assembly and maturation initiation | 113 |

| | | |
|-----|---|------------|
| 5.3 | Interaction of the capsid with host cofactors | 114 |
| 5.4 | Perspectives: immature lattice assembly, maturation initiation and capsid stability | 115 |
| | References | 117 |
| | Publications and Permissions | 139 |
| | Acknowledgements | 142 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | Map of major features in the HIV-1 RNA genome | 2 |
| 1.2 | Gag domains and proteolytic cleavage sites | 3 |
| 1.3 | Cartoon representation of the fold of immature HIV-1 CA | 4 |
| 1.4 | Schematic representation of the stages in the HIV-1 life cycle | 11 |
| 1.5 | Architectural maturation of HIV-1 viral particles visualised by cryo-electron tomography | 13 |
| 1.6 | Example of a one-dimensional contrast transfer function in focus and a $-1.5\ \mu\text{m}$ underfocus | 17 |
| 1.7 | Illustration of tilt series acquisition and reconstruction by weighted back-projection | 19 |
| 1.8 | Schematic illustrating the missing wedge in electron tomography | 20 |
| 1.9 | Workflow of a typical reference-free subtomogram alignment experiment | 22 |
| 2.1 | Schematic of the structural changes that occur during maturation of HIV-1 CA | 29 |
| 2.2 | Schematics and biochemical characterisation of the cleavage mutant constructs used | 32 |
| 2.3 | Distribution of phenotypes observed in each of the Gag cleavage mutant data sets | 41 |
| 2.4 | Lattice map representation of CA architecture in each cleavage mutant . . | 43 |
| 2.5 | Immature hexamer structures from Gag cleavage mutant viruses | 45 |

| | | |
|-----|--|----|
| 2.6 | Fit of the wild-type CA atomic model from immature Gag into immature cleavage mutant hexamer maps | 46 |
| 2.7 | Mature hexamer structures from Gag cleavage mutant viruses | 47 |
| 2.8 | Mature pentamer structures from Gag cleavage mutant viruses | 48 |
| 2.9 | Fourier shell correlation (FSC) for immature and mature CA hexamer and pentamer structures in all cleavage mutants | 49 |
| 3.1 | Possible combinations of missing neighbours in a hexagonal lattice | 57 |
| 3.2 | Central XY slice through the references constructed for multi-reference subtomogram classification | 61 |
| 3.3 | Computational slices through an example tomogram with overlaid subtomogram positions | 63 |
| 3.4 | Eigenvolumes selected for clustering of the subtomogram data set based on wedge-masked difference maps | 65 |
| 3.5 | Structures of incomplete hexamers at immature Gag lattice edges obtained by WMD PCA classification | 66 |
| 3.6 | Fourier shell correlation (FSC) of final WMD PCA incomplete hexamer classes | 67 |
| 3.7 | Side view of incomplete hexamer maps from the direction of the missing neighbouring hexamers | 68 |
| 3.8 | Schematics of the interactions stabilising the immature HIV-1 Gag lattice | 71 |
| 4.1 | Representative negative stain EM image of <i>in vitro</i> CA cores assembled with IP ₆ | 89 |
| 4.2 | Computational slices through a tomogram of conical, <i>in vitro</i> CA cores . . | 90 |
| 4.3 | Mature CA hexamer structure at 3.9 Å resolution from IP ₆ -stabilised <i>in vitro</i> cores | 91 |
| 4.4 | Mature CA pentamer structure at 6.2 Å resolution from IP ₆ -stabilised <i>in vitro</i> cores | 92 |

| | | |
|------|---|-----|
| 4.5 | Comparison of the density in the R18 pore of the hexamer in VLPs with IP ₆ | 94 |
| 4.6 | Lattice curvature classification of CA VLPs assembled with IP ₆ | 96 |
| 4.7 | Comparison of the density in the R18 pore of the pentamer in VLPs with IP ₆ | 97 |
| 4.8 | Fourier shell correlation (FSC) for structures obtained from CA VLPs with IP ₆ | 98 |
| 4.9 | Structure of the CA hexamer from VLPs incubated with 50 μM CPSF6 peptide (residues 313-327) | 99 |
| 4.10 | Binding pocket in the CA pentamer and adjacent hexamers with 50 μM CPSF6 peptide (residues 313-327) | 100 |
| 4.11 | Bivariate histogram of hexamer pair twist angle versus tilt angle in CA VLPs with CPSF6 peptide | 101 |
| 4.12 | Structure of the CA hexamer from VLPs incubated with 50 μM Nup153 peptide (residues 1407-1423) | 101 |
| 4.13 | Binding pocket in the CA pentamer and adjacent hexamers with 50 μM Nup153 peptide (residues 1407-1423) | 102 |
| 4.14 | Bivariate histogram of hexamer pair twist angle versus tilt angle in CA VLPs with Nup153 peptide | 103 |
| 4.15 | Fourier shell correlation (FSC) for CA structures with IP ₆ and 50 μM Nup153 or CPSF6 peptide | 104 |
| 4.16 | Variability in the side chain positions of the neighbouring CA molecule with curvature in the Nup153 binding site | 105 |

List of Tables

| | | |
|-----|--|----|
| 2.1 | Summary of data acquisition parameters for all HIV-1 Gag cleavage mutants | 34 |
| 2.2 | Summary of image processing parameters for all HIV-1 Gag cleavage mutants | 35 |
| 4.1 | Summary of data acquisition parameters for HIV-1 CA cores assembled with IP ₆ | 82 |
| 4.2 | Summary of image processing parameters for HIV-1 CA cores assembled with IP ₆ | 83 |

Abbreviations

| | |
|-----------------|---|
| 3D | Three-dimensional |
| AIDS | Acquired Immunodeficiency Syndrome |
| ALIX | ALG2-interaction protein X |
| APOBEC3G | Apolipoprotein B mRNA-editing catalytic polypeptide |
| BSA | Bovine serum albumin |
| CA | Capsid domain |
| CA-CTD | Capsid C-terminal domain |
| CA-NTD | Capsid N-terminal domain |
| CCC | Cross-correlation coefficient |
| CCD | Charge-coupled device |
| CG | Coarse-grained |
| CHMP | Charged multivesicular body protein |
| CLEM | Correlative light and electron microscopy |
| CPSF6 | Cleavage and polyadenylation specific factor 6 |
| Cryo-EM | Cryo-electron microscopy |
| Cryo-ET | Cryo-electron tomography |
| CTD | C-terminal domain |
| CTF | Contrast transfer function |
| CypA | Cyclophilin A |
| DMSO | Dimethyl sulfoxide |
| DQE | Detective quantum efficiency |

| | |
|-----------------------|--|
| EDTA | Ethylenediaminetetraacetic acid |
| EM | Electron microscopy |
| ESCRT | Endosomal Sorting Complex Required for Transport |
| FEZ1 | Fasciculation and elongation protein zeta 1 |
| FIB | Focused ion beam |
| FSC | Fourier shell correlation |
| Gag | Group-specific antigen |
| HAC | Hierarchical ascendant clustering |
| HIV-1 | Human immunodeficiency virus type 1 |
| IN | Integrase |
| IP₅ | Inositol pentakisphosphate |
| IP₆ | Inositol hexakisphosphate |
| LTR | Long terminal repeat |
| MA | Matrix domain |
| MD | Molecular dynamics |
| ML | Maximum likelihood |
| MSA | Multivariate statistical analysis |
| NC | Nucleocapsid domain |
| Nef | Negative factor |
| NMR | Nuclear magnetic resonance |
| NTD | N-terminal domain |
| Nup153 | Nucleoporin 153 |
| Nup358 | Nucleoporin 358 |
| PBS | Phosphate buffered saline |
| PCA | Principal component analysis |
| PEI | Polyethylenimine |
| PIC | Pre-integration complex |
| PIP2 | Phosphatidylinositol-(4,5)-bisphosphate |
| PR | Protease |

| | |
|-----------------|---|
| PSF | Point spread function |
| RANBP2 | RAN binding protein 2 |
| Rev | Regulator of expression of viral proteins |
| RNP | Ribonucleoprotein complex |
| ROI | Region of interest |
| RRE | Rev Response Element |
| RT | Reverse transcriptase |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SIV | Simian immunodeficiency virus |
| SNR | Signal-to-noise ratio |
| SP1 | Spacer peptide 1 |
| SP2 | Spacer peptide 2 |
| SVD | Singular value decomposition |
| TAR | Trans-activation response |
| Tat | Trans-activator of transcription |
| TCEP | Tris(2-carboxyethyl)phosphine |
| tICA | Time-structure based independent component analysis |
| UTR | Untranslated region |
| Vif | Viral infectivity factor |
| VLP | Virus-like particle |
| Vpr | Viral protein R |
| Vpu | Viral protein U |
| WMD | Wedge-masked difference map |
| WT | Wild-type |

Introduction

1.1 Human immunodeficiency virus type 1 (HIV-1)

Human immunodeficiency virus type 1 (HIV-1) is a lentivirus in the family *Retroviridae* and is the main causative agent of Acquired Immunodeficiency Syndrome (AIDS), a disease that was first recognised clinically in 1981 and which rapidly became a global pandemic (Greene, 2007). Both HIV-1 and the related lentivirus HIV-2 are thought to have originated from multiple transmissions of related simian immunodeficiency viruses (SIVs) between African primates and humans, possibly upon exposure to blood or mucosal fluids via the bushmeat trade (Sharp & Hahn, 2011). These multiple cross-species transmissions are reflected in the different viral lineages of HIV-1: M (main), N (non-M, non-O), O (outlier) and P (putative), of which the M group is the pandemic form (Sharp & Hahn, 2011).

HIV-1 infects CD4⁺ T-lymphocytes of the immune system, causing a gradual depletion of these cells over a period of asymptomatic latency that can last for years (Coffin & Swanstrom, 2013). As T-cell levels fall, eventually opportunistic infections can no longer be controlled by the immune system, leading to the onset of the symptoms of AIDS, including dementia, fever and wasting (Coffin & Swanstrom, 2013). Antiretroviral treatments are now available and have greatly reduced the mortality rate of HIV-1 infection, but significant challenges remain due to the evolution of drug-resistant HIV-1 strains that render current treatments ineffective (Collier et al., 2019). Research into the life cycle of HIV-1 infection is therefore of great importance in order to identify new potential drug targets that can be used as part of combination therapy to overcome the problem of resis-

tance as well as in longer-term efforts to develop a vaccine.

1.1.1 Organisation of the HIV-1 genome

In common with other retroviruses, HIV-1 possesses two identical copies of a positive-sense RNA genome that is reverse transcribed to produce a double-stranded proviral DNA molecule (also known as the provirus), which is integrated into the chromatin of the host cell in order to allow for expression of the various protein components of the virus required for viral replication (Wu, 2004). The unusual presence of two copies of the viral genome facilitates high rates of recombination between the two RNA strands (Hu & Temin, 1990), which provides a significant advantage to the virus by facilitating the evolution of resistance to antiretroviral drugs (Nora et al., 2007).

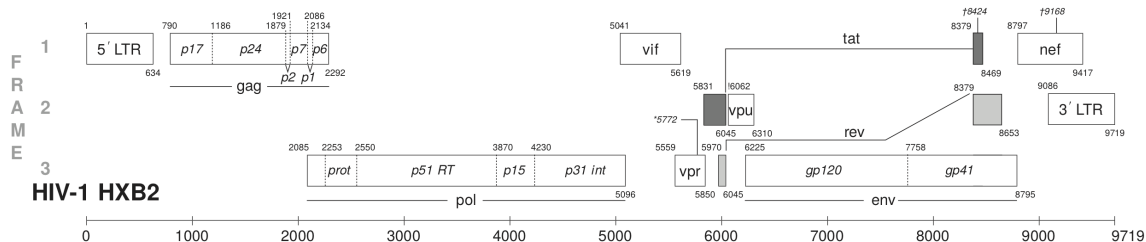


Figure 1.1: Map of major features in the HIV-1 genome. Adapted and reproduced with permission from Foley et al. (2018).

The approximately 9.7 kilobase HIV-1 RNA genome packaged into the virus is unspliced, 5'-capped and polyadenylated at its 3' end, containing nine genes that are flanked by 5' and 3' untranslated regions (UTRs), consisting of long terminal repeat (LTR) sequences (Lu et al., 2011). Figure 1.1 illustrates the major features of the HIV-1 genome. *gag*, *pol* and *env* are the genes that encode the main polypeptides of the virus, Gag (group specific antigen), Pol (polymerase) and Env (envelope) respectively.

The *gag* and *pol* genes lie in different reading frames with the start of *pol* overlapping with the 5' end of the *gag* at a U UAA sequence, and -1 ribosomal frameshifting at the UAA stop codon allowing translation of a leucine codon instead, resulting in addition of Pol at the end of Gag (Figure 1.1) (Jacks et al., 1988). This results in approximately 5 % of Gag being produced as the longer Gag-Pol polyprotein (Jacks et al., 1988). The genome also contains genes that encode the accessory proteins Vpr (viral protein R), Vpu (viral protein U), Vif (viral infectivity factor) and Nef (negative factor), as well as genes encoding the regulatory proteins Tat (trans-activator of transcription) and Rev (regulator of expression of viral proteins) (Wu, 2004). These viral proteins will be introduced in more detail in the next section.

1.1.2 HIV-1 proteins

Gag and Gag-Pol

The main structural component of the virus is Gag (group-specific antigen), a 55 kDa polypeptide containing the independently-folded domains MA (matrix), CA (capsid), NC (nucleocapsid), p6 and the spacer peptides SP1 and SP2, separated by flexible linkers (Figure 1.2A) (Briggs & Kräusslich, 2011). Gag undergoes proteolytic processing by the HIV protease during viral maturation, and this processing at the five cleavage sites frees its component domains in order to allow structural rearrangement (Mattei et al., 2016b) (Figure 1.2B). The roles of the various Gag domains, and in particular CA, will be discussed in more detail in following sections.

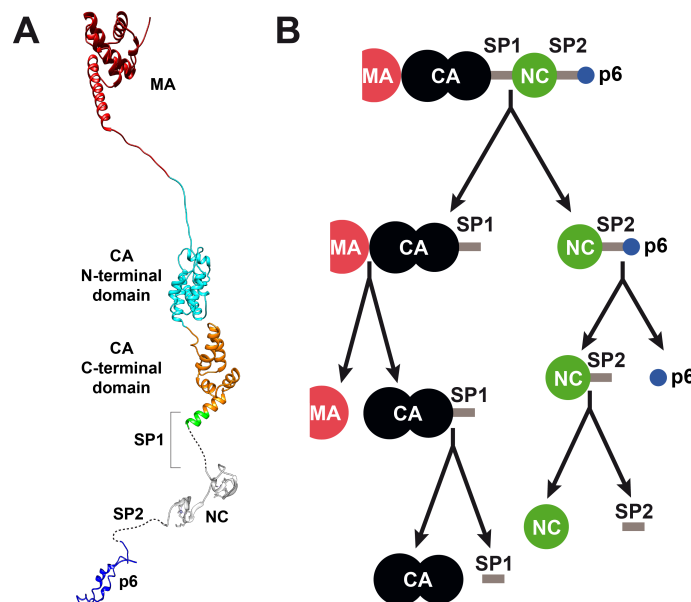


Figure 1.2: Gag domains and proteolytic cleavage sites **A** Cartoon representation of domains present in HIV-1 Gag. **B** Schematic of Gag proteolytic cleavages, arranged from top to bottom according to the relative rates of cleavage, with faster cleavages on top. Cleavage rates were determined by Pettit et al. (2005) on the cleavage site sequences in solution. Figure panel B was originally produced by myself for publication in Mattei et al. (2018), and has been reproduced here. See Publications and Permissions section for details.

The *pol* gene encodes proteins essential to the replication of the viral genome and its integration into the host cell genome, namely the viral protease (PR), reverse transcriptase (RT) and integrase (IN). As explained in the previous section, a ribosomal frameshift between the *gag* and *pol* genes results in production of Gag-Pol at about 5 % of total Gag expression, resulting in the incorporation of a 1:20 ratio of 160 kDa Gag-Pol to 55 kDa Gag in the virus (Jacks et al., 1988).

MA

The 17 kDa MA (matrix) domain is responsible for targeting Gag to lipid rafts, domains enriched in sphingomyelin and cholesterol on the plasma membrane of the host cell during viral assembly (Zhou et al., 1994). MA targeting to the membrane is achieved via a bipartite motif consisting of N-terminal myristoylation as well as a cluster of highly basic amino acids (Yuan et al., 1993; Zhou et al., 1994). In the monomeric form of MA, this myristyl group is sequestered within the protein and behaves like a switch, adopting a more exposed conformation upon MA trimerisation (Tang et al., 2004). The lipid phosphatidylinositol-(4,5)-bisphosphate (PIP2) plays a key role in this process (Ono et al., 2004) by binding to the MA protein and triggering the switch from the sequestered conformation to the more exposed conformation, which facilitates insertion of the myristyl group into the plasma membrane (Saad et al., 2006).

CA

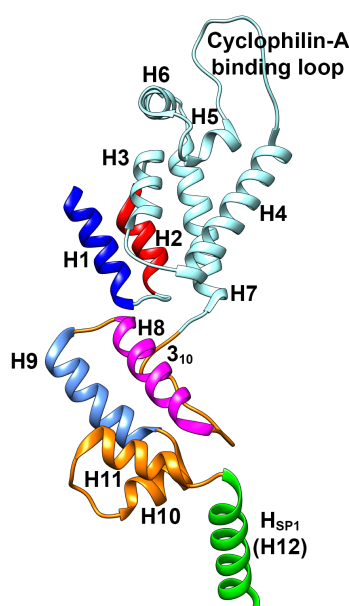


Figure 1.3: Cartoon representation of the fold of immature HIV-1 CA. Helices 1 to 11 of CA are denoted H1 to H11, and H12 is the CA-SP1 alpha helix formed between the C-terminus of CA and the N-terminus of SP1. The coordinates shown are from an atomic model built from a high-resolution cryo-EM map of the immature CA domain (PDB accession code: 5L93).

Downstream of MA is the 24 kDa CA domain, which consists of two main structural regions, the CA N-terminal domain (NTD) and the CA C-terminal domain (CTD) which both have an alpha-helical fold (Figure 1.3). Residues 1 to 145 form the NTD of CA, which in HIV-1 consists of 7 alpha helices with a loop between helices 4 and 5 that binds to the cellular cofactor cyclophilin A (CypA) (Gres et al., 2015).

The immature form of CA within Gag contains a long, flexible linker upstream of helix 1 that extends down from the MA domain and contains a proteolytic cleavage site. Proteolytic processing of this site allows the C-terminal end of this linker to fold down into a beta-hairpin structure (Gitti et al., 1996). Multiple roles have been suggested for this beta-hairpin in the virus, including as a structural switch in the maturation of CA (Gross et al., 1998; von Schwedler et al., 1998) and possible regulatory functions in the import of nucleotides into the mature capsid (Jacques et al., 2016). The question of the role of the beta-hairpin is examined in detail in Chapter 2.

Residues 146 to 149 form a flexible linker after which residues 150 to 231 comprise the CTD of CA, which consists of short 3_{10} helix and 4 alpha helices (Gres et al., 2015). An additional alpha helix is formed between the C-terminal end of CA and the first ten residues of the neighbouring SP1 peptide but is only present in the immature virus (Schur et al., 2016). This is explained in more detail in the next section

CA oligomerises to form lattices of CA hexamers, but these hexamers are very different between the immature and mature forms of the virus, as there is a large relative rotation between the NTD and CTD of CA between the immature and mature hexamer forms that gives rise to completely different intra- and inter-hexamer stabilising interactions (Mattei et al., 2016a; Schur et al., 2016). This is explained in more detail in the following sections.

SP1

SP1 (spacer peptide 1) is a stretch of 14 amino acids between the C-terminal end of the CA domain and the N-terminal end of NC which plays an important role in viral assembly, and deletion of this region from the Gag polyprotein completely abolishes infectivity (Kräusslich et al., 1995). The exact function of the SP1 domain was the subject of speculation for a long time, but in recent years, it was proven that N-terminal 10 residues of SP1 adopt a helical conformation and form an amphipathic helix together with the last 8 residues of the CA C-terminal domain if Gag dimerisation is stimulated *in vitro* by fusing it to a leucine zipper (Datta et al., 2016). Subsequently, Schur et al. (2016) determined the structure of the CA domain from the immature Gag lattice of authentic viruses as well as Gag virus-like particles (VLPs). These structures showed conclusively that the immature CA hexamer within the Gag lattice contains a 6-helix bundle motif formed at its C-terminal end together with the first part of SP1.

The cleavage site between CA and SP1 lies within this 6-helix bundle and it is cleavage at this site that is the slowest step in the proteolytic cleavage cascade involved in HIV-1 maturation (Pettit et al., 2005). This agrees with the well-known role of the CA-SP1 junction

in the regulation of proteolytic cleavage during the maturation process (Kräusslich et al., 1995; Wiegers et al., 1998).

It is therefore unsurprising that this 6-helix bundle is the target of a class of antiretroviral drugs known as maturation inhibitors, which stabilise the helical conformation of this junction (Schur et al., 2016; Wagner et al., 2016). Interestingly, resistance mutations to maturation inhibitors are not within binding pockets for these drugs, but are spread throughout the C-terminal domain of immature CA and destabilise the immature CA hexamer, thereby overcoming the increased stability of the 6-helix bundle effected by these drugs (Schur et al., 2016).

NC

Downstream of SP1 is the 6.4 kDa NC (nucleocapsid) domain of HIV-1, which is responsible for recognition and packaging of the viral RNA genome. In common with all other retroviral lineages, with the exception of the spumaviruses, the NC domain performs this RNA-binding function via a CCHC-type zinc finger domain (Guzman et al., 1998; Summers et al., 1990) that recognises four stem-loop structures formed by a sequence called ψ , located between the 5' UTR and the start of the *gag* open reading frame (Lever et al., 1989; Stephenson et al., 2013). In the immature virus, the NC domain plays an important role in viral assembly, and it has been shown that Gag polymerisation at the plasma membrane is facilitated by NC-RNA interactions that scaffold the nascent lattice (Pak et al., 2017; Yang et al., 2018).

In the mature virus, Gag cleavage results in the condensation of NC and RNA to form a dense region observed in electron micrographs termed the ribonucleoprotein complex (RNP) (Briggs & Kräusslich, 2011). This has been proposed to nucleate assembly of the mature capsid, but the presence of multiple capsids per virus in some viral preparations complicates this explanation (Briggs et al., 2003). Indeed, Mattei et al. (2015) showed that replacing the NC and SP2 domains of HIV-1 Gag with a leucine zipper still allows for the formation of capsids with wild-type morphology, although this did lead to a high proportion of aberrant capsids forming. This indicates that assembly of the mature capsid does not require NC or RNA condensation.

SP2

The 10-amino acid SP2 (spacer peptide 2) domain lies downstream of NC in HIV-1 Gag, and has been shown to be dispensable for maturation of the virus (de Marco et al., 2012).

Deletion of SP2 still allows the ribosomal frameshift between *gag* and *pol* to occur and is compatible with viral maturation, however the overall kinetics of proteolytic processing in the region between NC, SP2 and p6 still appears to be an important determinant of infectivity (de Marco et al., 2012).

p6

The final domain of Gag is p6, which plays an important role in recruiting the host cell Endosomal Sorting Complex Required for Transport (ESCRT) system to facilitate viral budding from the host cell. The PTAP motif within HIV-1 p6 recruits the Tsg101 subunit of the ESCRT-I system whereas its YPX_nL motif recruits ALG2-interaction protein X (ALIX) from the ESCRT-III system, both of which in turn recruit other ESCRT-III components from the charged multivesicular body protein (CHMP) family as well as the ATPase Vps4 to the budding site. These CHMP proteins polymerise into a dome-like structure in the viral bud and Vps4 then triggers membrane scission and release of an immature virus. These processes are reviewed in Sundquist and Kräusslich (2012) and Lippincott-Schwartz et al. (2017).

There is also evidence that HIV-1 budding can still occur in the absence of p6, presumably via an ESCRT-independent pathway, as premature cleavage between SP1 and NC which causes early detachment of the p6 domain still allows formation of budded immature viral particles, albeit with aberrant Gag lattice morphology (Carlson et al., 2010). Chapter 2 discusses this phenotype in more detail.

PR

The HIV-1 protease (PR) is an 11 kDa protein expressed as part of the Gag-Pol polyprotein, which comprises approximately 5 % of the total Gag protein within the virus (Graves et al., 1988; Jacks et al., 1988). PR is an aspartic protease with a fold consisting of mostly beta-sheets, and functions as a dimer with the catalytic site formed by one aspartate side chain from each monomer, held in place by a network of hydrogen bonds (Graves et al., 1988).

The activity of the HIV-1 protease is very low in its monomeric form and therefore dimerisation of the PR domains of adjacent Gag-Pol molecules is a prerequisite for maturation (Ishima et al., 2003). Protease activation upon Gag-Pol dimerisation occurs via an autocatalytic mechanism, in which the PR domain within Gag-Pol initially cleaves itself in order to form a mature dimeric protease that can process the rest of Gag and Gag-Pol

(Pettit et al., 2004). The HIV-1 protease is an actively-exploited pharmacological target in antiretroviral therapy, used by a class of drug known as protease inhibitors (Ghosh et al., 2016).

Mature protease dimers cleave Gag at 5 defined sites (Figure 1.2B), releasing its component domains which allows them to rearrange to form a mature virus with a closed capsid that encloses a condensed RNP genome. These cleavages occur at drastically different rates, with differences of up to 400-fold between the fastest and the slowest cleavages as determined in solution (Pettit et al., 2005). Proteolytic maturation of Gag is introduced in more detail in the next section and is discussed further in Chapter 2.

RT

Reverse transcriptase (RT) is one of the defining features of retroviruses, which use this enzyme to synthesise a DNA copy of their RNA genome during the viral life cycle (Hu & Hughes, 2012). RT is expressed as part of the Gag-Pol polyprotein and is a heterodimeric complex consisting of the p66 subunit, which contains a DNA polymerase domain as well as a ribonuclease H (RNase H) domain, and the p51 subunit which serves a structural function and is quite similar to the polymerase domain of p66 (Lightfoote et al., 1986).

The polymerase domain uses the host tRNA Lys3, which is complementary to the primer binding site near the 5' end of the viral RNA, in order to synthesise a negative strand DNA copy of the viral RNA genome (Hu & Hughes, 2012). The 5' and 3' UTR sequences flanking the viral genome are important in this process as the overhang they provide allows for strand annealing during DNA synthesis and copying of the entire genome (Hu & Hughes, 2012). The RNase H domain degrades the template RNA as synthesis progresses, but the 3' polypurine tract region in the viral RNA is resistant to this degradation and acts as the primer for synthesis of the positive DNA strand (Hu & Hughes, 2012).

Reverse transcriptase inhibitors form a key part of current antiretroviral therapy approaches (Hu & Hughes, 2012). There has also been interest in the role of reverse transcription in triggering HIV-1 capsid uncoating during infection. Atomic force microscopy experiments on isolated HIV-1 capsids suggest that reverse transcription causes pressure to build up inside the capsid, which may play a role in initiating uncoating (Rankovic et al., 2017). Uncoating also appears to depend on reverse transcription reaching at least the first strand transfer stage (Cosnefroy et al., 2016), and this reverse transcription-mediated uncoating can be blocked by compounds such as PF-74 that stabilise the capsid (Rankovic et al., 2018).

IN

The HIV-1 integrase (IN) is a 32 kDa protein that associates into multimeric assemblies with the proviral DNA known as intasomes for nuclear import and subsequent strand transfer (Delelis et al., 2008; Passos et al., 2017). Integrase mediates integration of the double-stranded DNA provirus into the genome of the host cell in a stable and permanent way, and this requires the pre-integration complex PIC, which includes components such as RT, IN and CA, to be imported across an intact nuclear membrane (Anderson & Maldarelli, 2018).

IN is actively targeted by antiretroviral drugs in clinical use, and aberrant capsids with a mislocalised RNP genome have been observed in studies with these drugs as well as with IN mutations that inhibit viral RNA binding, pointing to a role of IN in ensuring correct mature viral morphology (Anderson & Maldarelli, 2018). Targeting of integration is also modulated by capsid binders such as cleavage and polyadenylation specific factor 6 (CPSF6) (Achuthan et al., 2018), the binding of which to capsids will be discussed in Chapter 4.

Env

The *env* gene encodes the viral envelope glycoproteins responsible for mediating viral envelope fusion with the plasma membrane of a target cell. The precursor protein encoded by *env* is the transmembrane fusion protein gp160, which is expressed on the endoplasmic reticulum and undergoes glycosylation in the Golgi apparatus (Willey et al., 1988). gp160 trimerises and is cleaved by the host protease furin in the trans-Golgi, which produces mature glycoprotein spikes consisting of the proteins gp120 and gp41 (Hallenberger et al., 1992).

gp120 contains a glycosylated outer domain and a conserved inner domain that interacts with the gp41 protein, which in turn contains extracellular, cytoplasmic and transmembrane domains (Blumenthal et al., 2012). These spikes are incorporated into the viral envelope and mediate viral fusion via the receptor CD4 and co-receptors CCR5 and CXCR4 (Chen, 2019). Many details about the mechanism by which membrane fusion is mediated by HIV-1 Env are still unclear, but it is thought that gp120 binding to CD4 and CCR5 induces membrane curvature due to the rigidity of the CD4 ectodomain, and that gp120 dissociation from the Env trimer upon CD4 and CCR5 binding triggers structural reorganisation of the gp41 fusion fragment and insertion into the target cell membrane (Chen, 2019).

Regulatory and accessory proteins

The HIV-1 genome also encodes the regulatory proteins Tat and Rev, as well as a number of accessory proteins, namely Vpu, Vif, Vpr and Nef. These are briefly introduced here. Regulator of expression of viral proteins (Rev) is a regulatory protein that accumulates in the nucleus and oligomerises on a cis-acting viral RNA sequence known as the Rev Response Element (RRE), mediating export of viral transcripts from the nucleus via a nuclear export sequence that is regulated by transient masking via Rev multimerisation (Behrens et al., 2017). Trans-activator of transcription (Tat) is involved in regulating transcription of this RNA, by binding to a stem loop structure at the 5' end of the viral RNA known as the trans-activation response element (TAR element) and regulating the LTR promoter (Karn, 2011). A reduction in the level of expression or activity of Tat is associated with latency in HIV-1 infection, and this has led to interest in Tat as a target for antiretroviral therapy (reviewed in Karn (2011)).

Viral protein U (Vpu) is an 81-amino acid viroporin with a single transmembrane domain which regulates the distribution of membrane proteins on the host cell, and in this way, both enhances the release of immature HIV-1 particles during budding and down-regulates CD4 receptors after a cell is infected in order to evade detection by the host immune system (González, 2015). CD4 receptor downregulation is also mediated by another accessory protein, negative factor (Nef), which is crucial for replication of HIV-1 in primary T-lymphocytes (Lundquist et al., 2002).

Viral protein R (Vpr) is another accessory protein that performs several different functions during HIV-1 infection, such as targeting of the pre-integration complex (PIC) for nuclear import, acting as a co-activator of transcription at the LTR, regulating apoptosis of T-lymphocytes and inducing cell cycle arrest (Kogan & Rappaport, 2011). Viral infectivity factor (Vif) acts as a virulence factor which is essential for HIV-1 replication in host cells and enables the virus to overcome restriction by the host innate immune system, specifically by preventing mutagenesis of the viral genome mediated by apolipoprotein B mRNA-editing catalytic polypeptide (APOBEC3G) (Sharkey et al., 2019).

1.1.3 HIV-1 infection and replication

Lentiviruses are unique in infecting non-dividing cells (Novikova et al., 2019), which poses a problem to the virus as the proviral DNA must be transported across an intact nuclear membrane for genome integration to occur (Figure 1.4). The HIV-1 capsid, a closed protein shell encasing the viral ribonucleoprotein complex (RNP) and associated proteins, helps to overcome this problem (Freed, 2015). The capsid protects the viral RNA

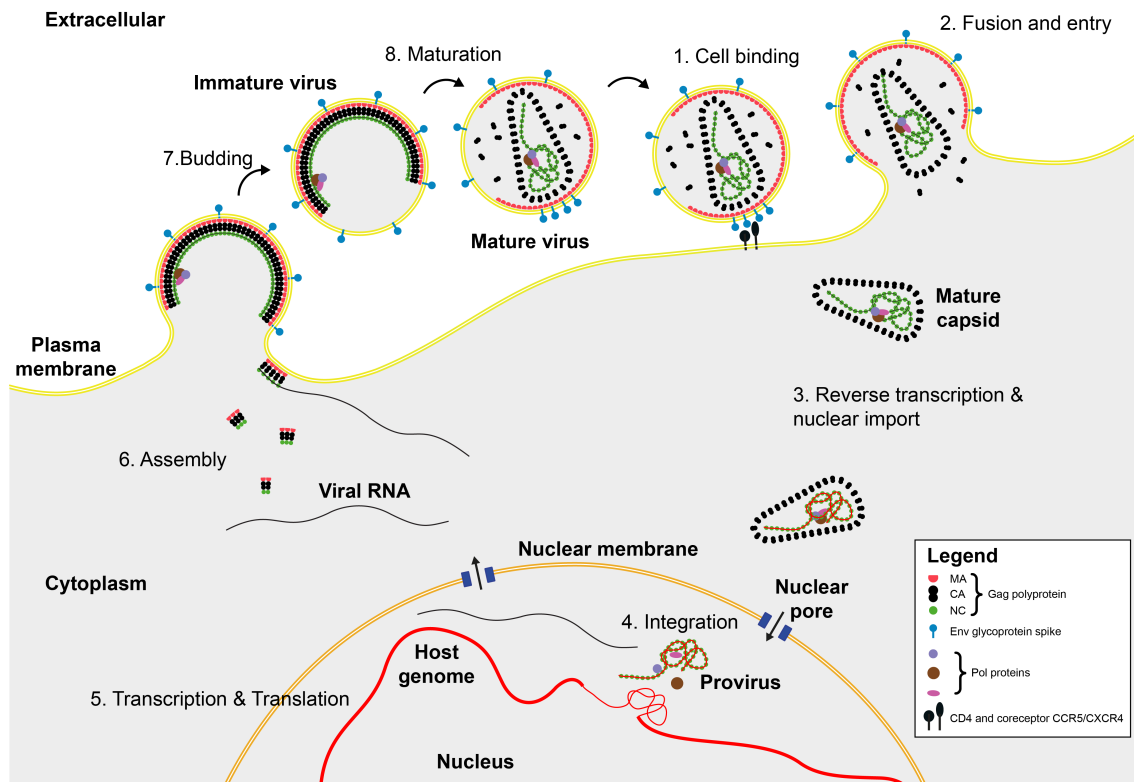


Figure 1.4: Schematic representation of the stages in the HIV-1 life cycle. Figure reproduced in adapted form with permission from Mattei (2016).

from degradation by cellular nucleases and mediates transport of the viral pre-integration complex (PIC) across nuclear pores (Freed, 2015). The key role of the HIV-1 capsid in various stages of the viral life cycle has therefore made it a very attractive target for development of new antiretroviral compounds as well as optimisation of existing drugs (Tedbury & Freed, 2015).

1.1.4 HIV-1 assembly

Transcription of proviral DNA after integration into the host cell genome results in the production of the viral mRNA that serves both as the form in which the viral genome is encapsulated in the virus, as well as being translated on polysomes to produce Gag and Gag-Pol (Sundquist & Kräusslich, 2012). Once translated, Gag is trafficked to nascent sites of viral assembly at the plasma membrane of the infected cell and recruits other viral components to the assembly site. The basic unit of of lattice assembly is still inconclusive, but it has been suggested to be either a dimer or trimer of Gag (Crist et al., 2009). Chapter 3 examines this question in more detail and presents evidence that sheds light on the identity of this basic unit of lattice assembly.

Gag associates with the plasma membrane at the assembly site via its MA domain, which

is anchored to the membrane via a myristoyl tail (Gheysen et al., 1989), and recruits the viral envelope glycoproteins (Env) (Yu et al., 1993). The viral RNA is recruited by the NC domain and plays an important role in scaffolding the assembly of the immature Gag lattice and promoting its growth (Pak et al., 2017; Yang et al., 2018). The CA domain of Gag is mainly responsible for driving assembly of the immature Gag lattice via formation of immature CA hexamers, which tile into a hexagonal lattice under the membrane (Briggs & Kräusslich, 2011). In order to accommodate the curvature of the membrane induced by MA, the immature CA hexameric lattice incorporates irregular defects at which no Gag hexamers are present, as well as a single, large gap at the budding site, facing the inside of the cell (Briggs et al., 2009).

The small molecule inositol hexakisphosphate (IP₆) plays an important role in the assembly process, as shown by experiments on *in vitro* assemblies of Gag, which differ in assembly properties from native immature viruses unless IP₆ is added during assembly (Campbell et al., 2001). Recently, it was shown by Dick et al. (2018) that IP₆ acts as an assembly cofactor for HIV-1 Gag by facilitating the formation of the 6-helix bundle between the C-terminus of CA and the first 10 residues of SP1.

Budding of the assembled virus is triggered by recruitment of the host ESCRT system via the p6 domain of Gag (Sundquist & Kräusslich, 2012). Gag assembly and budding are competitive processes, which introduces an element of stochasticity to these processes, in that the growth of the assembling Gag lattice competes with the assembly of ESCRT-III rings that mediate membrane constriction and fission at the budding site (Carlson et al., 2010). The kinetics of assembly, budding and maturation appear to be coupled, and defects at any stage of these processes can result in non-infectious viruses (Bendjennat & Saffarian, 2016; Carlson et al., 2010).

1.1.5 HIV-1 maturation

Immediately after budding, the virus exists in its immature form, which is non-infectious unless it undergoes a process called maturation (Pornillos & Ganser-Pornillos, 2019). HIV-1 maturation encompasses a set of related processes, and thus three distinct types of maturation can be defined: proteolytic, structural and architectural. Proteolytic maturation is defined as the cleavage of Gag at its five proteolytic cleavage sites, by the viral protease, which releases the component domains of Gag (Figure 1.2B). Proteolytic maturation is a prerequisite for structural maturation, which is the structural rearrangement and breaking of stabilising interactions in the immature CA hexameric lattice with 8 nm spacing, to form a mature CA hexameric lattice with 10 nm spacing via different contacts (Mattei et al., 2016a; Schur et al., 2016). The dependence of structural maturation on

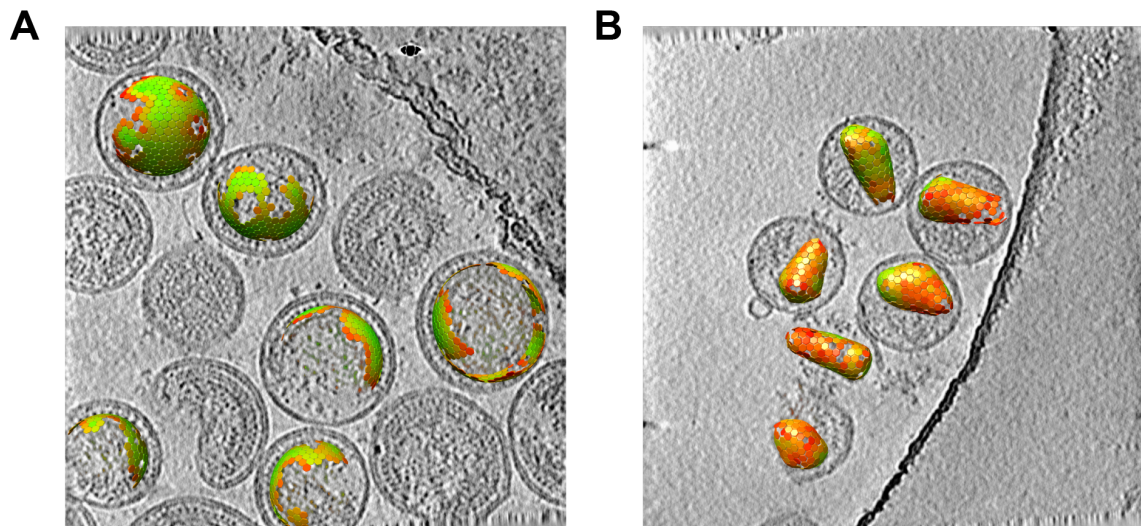


Figure 1.5: Architectural maturation of HIV-1 viral particles visualised by cryo-electron tomography. **A** Computational slice through a tomogram of purified immature HIV-1 viral particles, with overlaid lattice maps representing the positions and orientations of aligned subtomograms in order to visualise the morphology of the CA protein layer. The CA layer of Gag forms an incomplete, spherical shell under the viral envelope and incorporates small defects throughout as well as a single, large defect. Figure panel made by me using data acquired, processed and published by Schur et al. (2016). **B** As in A, but for a data set of purified mature HIV-1 viral particles. The CA protein in the mature virus forms closed, conical capsids by hexagonal tiling, although many capsids in purified viral samples have been observed to be defective or incomplete (Mattei et al., 2016b). Pentagonal gaps can be seen in the hexagonal lattice where CA pentamers are incorporated. Figure panel made by me using data acquired, processed and published by Mattei et al. (2016b).

the different cleavages that occur during proteolytic maturation is examined in detail in Chapter 2.

Architectural maturation describes the morphological changes in the CA protein assemblies present in the virus, which involve rearrangement of a truncated spherical immature lattice, containing defects and a large gap, into a closed, conical capsid that encapsulates the viral ribonucleoprotein (RNP) (Figure 1.5). Structural and architectural maturation require almost all of the interactions formed by CA in the immature lattice to be broken and re-formed due to the different lattice contacts involved in building a mature lattice (Mattei et al., 2016a; Schur et al., 2016). Although some have proposed a model in which the immature lattice rearranges in-place and folds up into a mature lattice (Frank et al., 2015), the resulting capsids would have to contain a seam or defect to accommodate this model, which is inconsistent with direct morphological evidence from subtomogram averaging of native viral capsids (Mattei et al., 2016a). Instead, most of the evidence in the literature is consistent with a model in which the immature lattice disassembles into lower-order oligomeric states of CA, and then reassembles to form a mature lattice, although the spatial and temporal details of this process are still unclear (Mattei et al., 2016b).

The immature lattice typically contains about 2400 ± 700 copies of Gag (Carlson et

al., 2008), but maturation involves both a reduction in diameter and increase in lattice spacing, and therefore it is not surprising that only between 1100 and 1300 copies of CA are actually incorporated into the mature capsid (Mattei et al., 2016a). If the disassembly-reassembly model for maturation is correct, the remaining copies of CA that have not been incorporated into the lattice must therefore exist freely in the lumen of the virus, possibly as monomers or oligomers. This free pool of CA in the virus is thought to exert a stabilising effect on the mature lattice before fusion with the target cell, as isolated capsids rapidly disassemble once in the cytosol (Márquez et al., 2018).

1.1.6 HIV-1 fusion and nuclear entry

Mature HIV-1 recognises CD4⁺ T-lymphocytes via the gp41/gp120 Env glycoprotein spikes on its surface, which bind to the cell via the CD4 receptor and co-receptors CCR5 and CXCR4, mediating membrane fusion and release of the contents of the virus into the cell (Permanyer et al., 2010). The capsid is then trafficked towards the nucleus and serves to protect the viral genome, preventing access by host nucleases that would otherwise degrade it (Freed, 2015). The capsid is also thought to serve as a vessel for reverse transcription, with the N-terminal beta-hairpin of the CA hexamer proposed to act as a pore for import of nucleotides (Jacques et al., 2016). Upon reaching the nucleus, the pre-integration complex (PIC) has to be imported across the nuclear pore to facilitate integration of the products of reverse transcription into the host genome. However, many details of this process are still unclear and it is not known to what extent capsid disassembly occurs at this stage.

These processes are modulated by binding of various small molecule and protein host cofactors to the capsid at various parts of the intracellular stage of its life cycle, a topic that has been of great interest in recent years. The surface of the capsid presents a variety of potential interfaces for interaction with cofactors within the host cell, and this binding modulates the mechanical stability, cellular trafficking and nuclear entry of the capsid. IP₆ in particular has been the subject of much attention, as it was recently reported by Mallery et al. (2018) that it is a pocket factor that binds to a pore formed by six arginine residues (R18) below the beta-hairpin in the mature capsid. Mallery et al. (2018) reported that IP₆ greatly increases the stability of HIV-1 capsids and prevents the spontaneous disassembly that would otherwise occur after viral fusion.

The capsid also interacts with various protein cofactors in the host cell, including cyclophilin A (CypA) (Franke et al., 1994), fasciculation and elongation protein zeta 1 (FEZ1) (Malikov et al., 2015), cleavage and polyadenylation specific factor 6 (CPSF6) (Lee et al., 2010), nucleoporin 153 (Nup153) and nucleoporin 358 (Nup358) (Di Nunzio

et al., 2012). CypA binds to a the cyclophilin binding loop between helices 4 and 5 of the CA-NTD, and has been shown to stabilise the capsid by changing its stiffness (Liu et al., 2016). FEZ1, a kinesin adaptor protein, was recently shown to bind to CA hexamers and mediate trafficking of the capsid along microtubules (Huang et al., 2019).

Binding of the capsid to nuclear pore components such as Nup153 and Nup358, as well as nuclear-localised cofactors such as CPSF6, plays an important role in targeting the capsid to the nuclear pore and mediating its entry. In particular, CPSF6 and Nup153 have been shown to be important in mediating translocation of the capsid and its contents into the nucleus (Bejarano et al., 2019; Buffone et al., 2018). These two cofactors bind to a common pocket formed between two adjacent CA molecules in the mature hexamer but the exact binding interface differs between the two (Price et al., 2014). The CPSF6/Nup153 binding pocket is also of pharmacological interest, as a number of small molecules such as PF74 and BI-2 have been found to bind with high affinity to this site and may modulate the stability and integrity of the capsid (Márquez et al., 2018). Binding of capsids to both IP₆ and the host cofactors CPSF6 and Nup153 are examined in more detail in Chapter 4.

1.2 Cryo-electron tomography (cryo-ET) and subtomogram averaging

1.2.1 Electron microscopy of vitrified biological samples

Vitrification of biological samples

The elucidation of many mechanistic details of the HIV-1 life cycle explained in the previous sections was made possible due to the availability of high-resolution structures of the molecular complexes in question that enabled protein and nucleic acid interactions to be mapped. Classically, X-ray crystallography has been the technique of choice for high-resolution structural determination of biological macromolecules, while nuclear magnetic resonance (NMR) spectroscopy has been the tool of choice to study the dynamics of such complexes (Bai et al., 2015). However, techniques such as X-ray crystallography and NMR spectroscopy require purified proteins or complexes at high concentrations, which removes them from their biological context within the organism being studied. In the case of X-ray crystallography, the additional requirement to form crystals of the protein or complex of interest can also lead to either crystal packing artefacts or conformations quite different from those found within their native environment. A relevant example of this is the HIV-1 CA pentamer within the mature capsid, which forms using different

protein-protein interactions within capsids in the native viral environment (Mattei et al., 2016a) compared to those observed when disulfide cross-linking is used to stabilise a pentamer for crystallisation (Pornillos et al., 2011).

Electron microscopy has become a powerful technique for the visualisation of biological samples, especially within the larger biological context in which they exist. Traditionally, the preparation of biological samples for imaging by electron microscopy involved staining with heavy metal salts (Brenner & Horne, 1959) or dehydration and plastic embedding of the sample (McEwen & Marko, 2001). However, the development of cryo-electron microscopy (cryo-EM) has enabled the imaging of biological macromolecules in a more native state by freezing them in a layer of vitreous ice Adrian et al. (1984), and this technique has undergone a ‘resolution revolution’ over the past decade with structural determination of these molecules to near-atomic resolution now possible (Kühlbrandt, 2014).

Vitrification is usually achieved by spreading a solution of the sample of interest onto a holey, amorphous carbon film suspended on a copper, molybdenum or gold support grid for electron microscopy. The solution is applied to the grid and blotted to leave a thin film which is then immediately plunged into liquid ethane to in order to freeze it without the formation of crystalline ice. The sample can then be imaged in a transmission electron microscope in a way that preserves its macromolecular structure by avoiding the use of stains or dehydration and retaining the aqueous environment in which it exists. Methods for preparation of biological samples for high-resolution structural determination by cryo-EM are reviewed in detail in Passmore and Russo (2016).

Imaging biological samples in the electron microscope

Cryo-EM imaging of biological samples is usually not done at focus, as biological macromolecules consist of relatively light elements that only scatter the incident electron beam weakly, leading to very little amplitude contrast in the resulting images at focus (Glaeser, 2013). Contrast in cryo-EM images is instead achieved by introducing a slight defocus to the objective lens of the microscope, which takes advantage of the weak phase object approximation to introduce phase contrast. Scattering of the incident electron beam by the sample introduces a phase shift in the scattered electron wave, which interferes with the unscattered electrons that pass through the sample, giving rise to phase contrast (Erickson & Klug, 1971). However, imaging away from optical focus also introduces a point spread function (PSF) which delocalises information in the resulting images and negatively affects high-resolution information (Erickson & Klug, 1971). The equivalent of the PSF in reciprocal space is known as the contrast transfer function (CTF), defined for each spatial

frequency f as

$$CTF(f) = A(\sin(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s))) + B \cos(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s))$$

where λ is the wavelength of the electron beam, c_s is the coefficient of spherical aberration, Δz is the defocus, A is the defocus-dependent envelope function that dampens the signal and B is the fraction of total contrast attributable to amplitude contrast (Erickson & Klug, 1971; Schur et al., 2013; Wade, 1992).

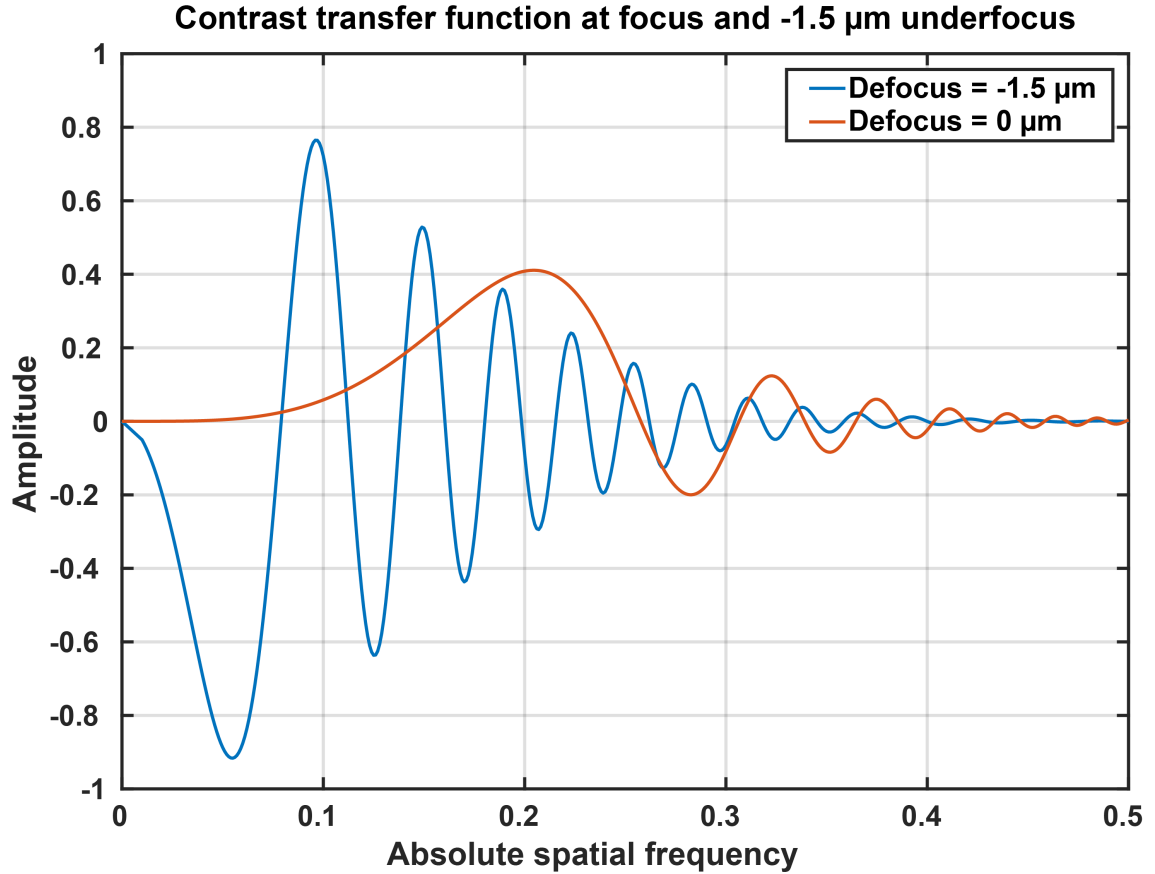


Figure 1.6: Example of a one-dimensional contrast transfer function in focus and a $-1.5 \mu\text{m}$ underfocus. The phase contrast transfer functions plotted here were simulated for a 300 keV Thermo Fisher Titan Krios microscope in MATLAB. Both curves are additionally dampened by an exponential at increasing spatial frequencies to represent the defocus-dependent envelope function present in actual images. The red curve shows that at focus, there is essentially zero contrast at low spatial frequencies, whereas using a defocus of $-1.5 \mu\text{m}$ drastically improves the low-resolution contrast in the image. However, it can also be seen that this results in contrast reversal at certain spatial frequencies in the image, which must be corrected for in order to restore high-resolution information.

It can be seen that the sinusoidal nature of the CTF when applied to the image results in contrast reversal at certain spatial frequencies (Figure 1.6), and the CTF must therefore be corrected for in order to restore high-resolution information in the image (Erickson & Klug, 1971). This is typically done by analysing power spectra of the images to obtain a more precise estimate of the defocus at which the sample was imaged than can be obtained from the microscope stage position, and then inverting the phase of the appropriate

regions of the CTF (Xiong et al., 2009). For thicker samples such as those typically imaged by cryo-electron tomography (explained in the following sections) the relative height difference between the top and the bottom of the sample, relative to the optical axis of the microscope, gives rise to a significant defocus gradient across the sample (Jensen & Kornberg, 2000). In such cases, an approach known as three-dimensional (3D) CTF correction can be used to take this defocus gradient across the sample into account by correcting for the CTF differently in stripes across the height of the sample (Jensen & Kornberg, 2000; Turoňová et al., 2017; Voortman et al., 2012). This has been shown to improve the resolution obtainable by subtomogram averaging of the immature HIV-1 Gag hexamer (Schur et al., 2016; Turoňová et al., 2017).

Ensemble techniques for structural determination

Another problem with using electron microscopy to image biological samples prepared in such a way is that biological molecules are very sensitive to radiation damage, which destroys high-resolution information by breaking chemical bonds in the sample (Glaeser, 2016). Although the imaging of the sample at liquid nitrogen temperatures slows this effect somewhat, the total electron dose that can be used for imaging is still quite low and thus the signal-to-noise ratio (SNR) of the resulting images is very poor (Meents et al., 2010; Taylor & Glaeser, 1976). The information of interest about high-resolution protein and nucleic acid structure is therefore far below the noise level, and computational techniques need to be employed in order to recover it. Ensemble averaging techniques such as single particle analysis overcome this problem by taking advantage of multiple, identical copies of the same molecule or complex in the sample, which exist in random orientations.

At sample thicknesses typically used for cryo-EM, these multiple particle views can be approximated as projection images, and according to the central slice theorem, their Fourier transforms are equivalent to central slices in the direction of projection through the three-dimensional Fourier transform of the object of interest (Crowther et al., 1970; DeRosier & Klug, 1968). Initially, the orientation of each of these particles is unknown, but this can be found by iterative alignment methods involving projection matching (Harauz & Van Heel, 1985) or maximum likelihood (ML) (Scheres, 2012) approaches, which then allows the projection images to be used in reconstructing an average three-dimensional structure of the particle (Crowther et al., 1970; DeRosier & Klug, 1968). Assuming a constant SNR and random noise in the projection images, averaging them leads to an amplification of the signal component relative to uncorrelated shot noise, recovering high-resolution information in the final 3D structure.

1.2.2 Tomography and three-dimensional reconstruction from projections

Both immature and mature HIV-1 particles are highly pleomorphic, exhibiting considerable variability in size and shape, and as such no two viral particles are completely identical (Briggs et al., 2003; Carlson et al., 2008). Even if only the CA protein layer is taken into consideration, the stochastic nature of immature lattice assembly and budding, as well as mature capsid assembly, means that the exact arrangement of CA hexamers in the immature and mature lattices in each virus is unique (Briggs et al., 2009; Briggs et al., 2003). This means that structural characterisation of HIV-1 viral particles by cryo-EM is not amenable to single particle analysis, unlike many other viruses which exhibit very high levels of structural homogeneity and helical or icosahedral symmetry (Caspar & Klug, 1962), and some of which such as adenovirus can even be crystallised to allow atomic-resolution structural determination (Reddy et al., 2010a; Reddy et al., 2010b). One powerful method that can be used to overcome this problem is cryo-electron tomography (cryo-ET).

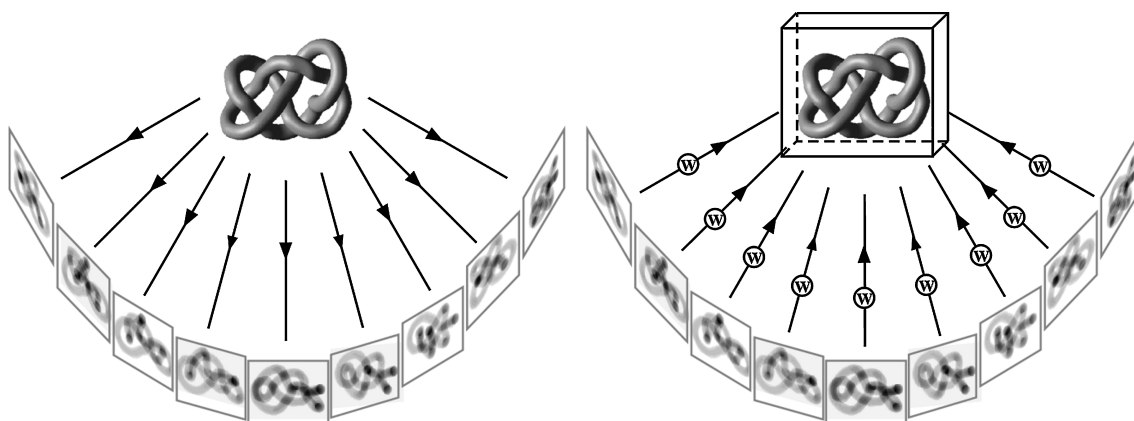


Figure 1.7: Illustration of tilt series acquisition and reconstruction by weighted back-projection. Figure from Lučić et al. (2005) and reproduced with permission of Annual Reviews in the format Thesis/Dissertation via Copyright Clearance Center.

Tomography involves physically tilting the sample on the stage of the electron microscope in defined angular increments across a range of tilt angles, and acquiring a two-dimensional projection of the sample at each tilt angle to collectively form a series of images called a tilt series (Figure 1.7). Acquiring a tilt series allows the three-dimensional structure of the volume to be reconstructed by effectively filling in different Fourier planes in reciprocal space of the three-dimensional volume (Crowther et al., 1970), producing a reconstruction known as a tomogram. In practice, tomogram reconstruction for high-resolution structural analysis is often done using a real space procedure known as weighted back-projection (Figure 1.7) (Obr & Schur, 2019). This involves back-projecting the real space tilt images into the three-dimensional tomogram volume while

applying a ramp filter in reciprocal space to compensate for the fact that the separation of the Fourier planes increases with increasing spatial resolution (Radermacher, 1992).

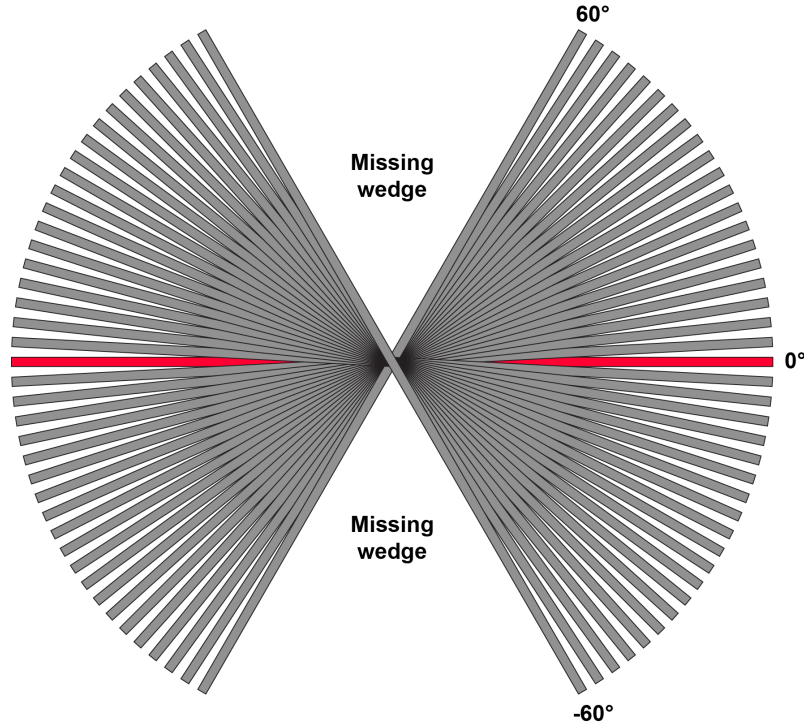


Figure 1.8: Schematic illustrating the missing wedge in electron tomography. The Fourier plane corresponding to the Fourier transform of the image at zero tilt is indicated in red, with Fourier planes corresponding to tilts in 3° increments across a $\pm 60^\circ$ range indicated in grey. The direction of the missing wedge of information in Fourier space due to physical tilt angle limitations is indicated. Note that the separation of the Fourier planes increases as distance from the centre of the reciprocal space image, corresponding to increasing spatial frequency, increases. This results in more overlap between the Fourier planes at lower spatial frequency and thus over-representation of low frequencies during reconstruction by back-projection, unless this is compensated for with a spatial frequency-dependent filter to correctly weight the different spatial frequencies.

In single particle analysis, the object of interest is present in multiple, identical copies and as long as enough projection views can be obtained of the object from different directions, Fourier space can be uniformly sampled for the reconstructed object and this allows for isotropic resolution in all directions. In contrast, tomography is normally used to image unique objects that do not exist as exactly identical copies of each other, and thus the projection views obtainable are limited by the physical geometry of the microscope stage which cannot tilt past a certain angle (usually $\pm 70^\circ$). Tilting the sample also makes it effectively thicker to the electron beam, according to the relationship

$$\text{Apparent thickness} = \frac{\text{Thickness at zero tilt}}{\sin(90^\circ - \theta)}$$

where θ is the tilt angle of the stage. The mean free path of 300 keV electrons through a vitreous ice is approximately 332 nm (Rice et al., 2018), with the apparent sample thickness increasing two-fold at a tilt angle of $\pm 60^\circ$. As the sample thickness approaches and

exceeds this mean free path, more inelastic scattering and multiple scattering events occur and contribute to loss of contrast in the images (Peet et al., 2019). These tilt angle limitations mean that Fourier space is incompletely sampled in the reconstructed tomograms, as the information corresponding to the reciprocal space planes outside the tilt range used (usually $\pm 60^\circ$) is missing, giving rise to what is termed the missing wedge problem. However, it is still possible to obtain high-resolution, isotropic structures of macromolecular complexes by cryo-electron tomography using a technique called subtomogram averaging, which is explained in the following section.

1.2.3 Subtomogram averaging

Subtomogram averaging is a method that is very useful in recovering high-resolution information from tomograms of biological samples that are too irregular for single particle analysis approaches. The principle of subtomogram averaging relies on the fact that, while different copies of objects such as the immature Gag lattice or the mature capsid of HIV-1 may be unique as a whole, they contain smaller features that are identical copies of each other, which in this example are immature and mature CA hexamers, respectively, that tile into a hexagonal lattice. Smaller subvolumes called subtomograms can be extracted from the tomograms which contain these features, and the coordinates at which subtomograms are to be extracted can be defined in a number of ways: manual picking, seeding a grid of points along a given geometry such as the surface of a sphere or tube or by segmentation of features such as membranes or viral capsids from within tomograms (Wan & Briggs, 2016).

In ‘reference-free’ approaches, these subtomograms are then averaged to produce an initial approximation of the structure that is relatively unfeatured, although the initial orientations of the subtomograms can usually be approximated quite well from the larger context of the tomogram (Wan & Briggs, 2016). The subtomograms are then iteratively aligned against the reference in a process analogous to alignment of two-dimensional particle images in single particle analysis, to refine the initial estimates of both the extraction coordinates and the Euler angles that define the orientation of the object of interest in each subtomogram (Figure 1.9) (Förster et al., 2005). Although the individual subtomograms are affected by the missing wedge problem as described in the previous section, a structure with isotropic resolution can still be obtained by averaging subtomograms that are rotated in different orientations relative to the optical axis of the microscope, thus filling in the missing information in Fourier space (Wan & Briggs, 2016).

Due to technical improvements such as the introduction of direct electron detectors with increased detective quantum efficiency DQE combined with the use of energy filters, as

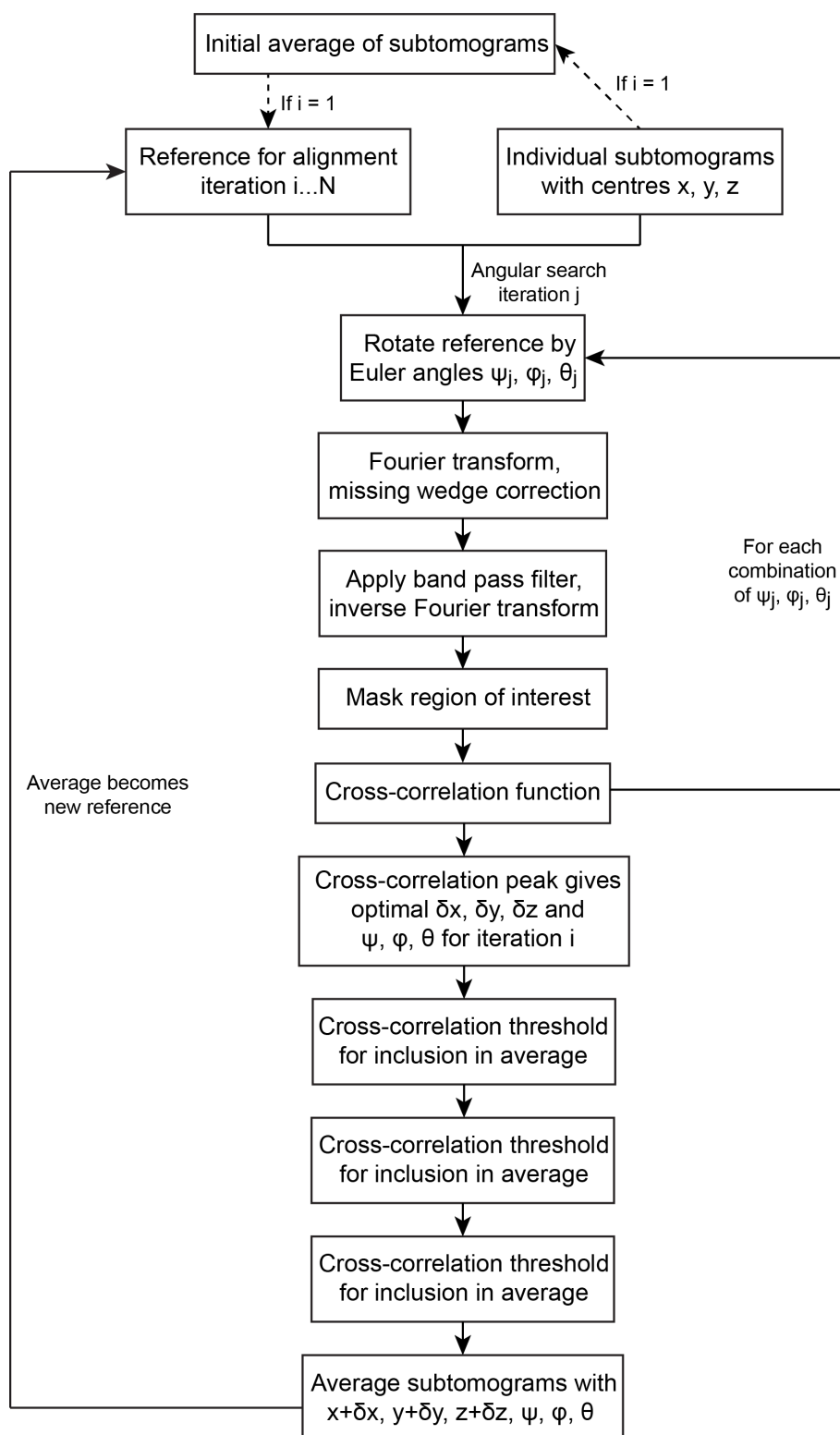


Figure 1.9: Workflow of a typical reference-free subtomogram alignment experiment, in which an unfeatured initial average of the subtomograms is used as a starting point to iteratively search for optimal box centre coordinates x , y and z , as well as the Euler angles ψ , θ and ϕ . Conventions for these Euler angles differ according to the software package used. This figure is based on the steps depicted in a subtomogram averaging schematic by Förster et al. (2005).

well as improvements in image processing approaches such dose-dependent frequency weighting and weighting (Grant & Grigorieff, 2015) and 3D CTF correction (Turoňová

et al., 2017), the resolution of structures attainable by subtomogram averaging has greatly improved over the past half a decade. This has led to the determination of the structure of the mature HIV-1 CA hexamer and pentamer to sub-nanometer resolutions (Mattei et al., 2016a) and of the immature HIV-1 Gag hexamer to near-atomic resolution (Schur et al., 2016). Subtomogram averaging is also very powerful because it can be used to determine structures of complex assemblies within their cellular context, by applying it to tomograms of cells that have been thinned by focused ion beam milling, without requiring biochemical purification that could perturb the system of interest (Bykov et al., 2017; Engel et al., 2015; Villa et al., 2013).

1.2.4 Image classification in cryo-EM

Classification is a powerful tool that can be used to separate images of biological macromolecules based on conformational differences. Image classification is integral to structural determination by single particle analysis, as it is essential in separating different particle views in projections and is also used to distinguish between different conformational states in 3D (Frank, 2006). One of the main problems in classification of cryo-EM data is the very low signal-to-noise ratio of the data, which stems from the fact that biological samples are very sensitive to radiation damage and therefore must be imaged with very low electron fluence in order to preserve the structure of the sample and thus high-resolution information (Glaeser, 2016). Different approaches have been developed in order to solve the problem of classifying cryo-EM images based on meaningful biological differences despite the high noise level in the data, including multivariate statistical analysis MSA (Borland & Heel, 1990; van Heel et al., 2016) and maximum likelihood (ML) based methods (Scheres, 2012; Scheres et al., 2007). MSA approaches will be introduced in more detail in the next section, including the classification algorithm that I implemented in the Briggs group subtomogram pipeline as explained in Chapter 3.

The classification methods used in single particle analysis can also be applied to subtomogram data sets, but this introduces additional challenges due to the nature of tomographic data. Of these, the missing wedge is the principal problem in subtomogram classification, as each subtomogram missing information in different parts of Fourier space due to its rotation with respect to the optical axis of the microscope. Combined with the noisy nature biological cryo-EM data, most of the variance within a data set of subtomograms is therefore due to the orientation of each subtomogram with respect to the missing wedge of its parent tomogram (Förster et al., 2008). Hence, in order to classify subtomograms according to actual differences in the biological object of interest, it is necessary to compensate for the missing wedge as described in the following sections.

Multivariate statistical analysis (MSA)

In both single particle analysis and subtomogram averaging experiments, a data set of n images consisting of p pixels each can be represented as the $n \times p$ matrix U , or as the equivalent $p \times n$ matrix U^T , such that each image corresponds to a pixel vector with grey-values defining its coordinates in p -dimensional space. This very high dimensionality combined with the low SNR makes classification of the images into distinct clusters difficult unless some method of dimensionality reduction is used. Multivariate statistical analysis methods have a long history in the field of biological electron microscopy and have been used quite successfully to reduce the dimensionality of this clustering problem (Borland & Heel, 1990; van Heel & Frank, 1981). This family of methods generally relies on eigendecomposition of a matrix of high dimensionality in order to reduce the representation of variance within the data set to a few eigenvectors, corresponding to only significant differences between the images (van Heel et al., 2016).

Correspondence analysis was the first MSA method to be applied in biological electron microscopy, and relies on computing a symmetric matrix of χ^2 distances between any two rows or columns of the pixel matrix U in order to compare relative differences between the data, followed by determination of the eigenvectors and eigenvalues of this matrix (van Heel & Frank, 1981). Projecting the image data set onto the new coordinate system defined by a small subset of chosen eigenvectors then reveals only the most significant differences between the images, which can then be divided into clusters along the new axes to divide the data set into classes (van Heel & Frank, 1981). However, computation of this χ^2 distance matrix requires only positive grey values in the images, which holds true for single particle projection images but is not the case for subtomograms due to the weighted back-projection algorithm often used to reconstruct tomograms (Förster et al., 2008; Radermacher, 1992). A related technique described by Borland and Heel (1990) solves this problem by replacing the χ^2 metric with a normalised metric with similar properties known as the modulation distance.

Other commonly-used MSA approaches involve calculating a covariance matrix or another matrix with a similarity measure such as cross-correlation, and then using singular value decomposition (SVD) to find the eigenvectors and corresponding eigencoefficients of the matrix (Frank, 2006). The matrix U representing the image data set is then projected onto the chosen eigenvectors to produce eigenimages, which are used to choose the eigenvectors to use as axes for clustering of the data into classes (Frank, 2006). Clustering of eigencoefficients can be achieved using different approaches, with k -means clustering and hierarchical ascendant clustering (HAC) being commonly used (Frank, 2006).

MSA methods have to be modified when applied to subtomogram classification in or-

der to account for the missing wedge. One approach that has become common in the field is to use constrained cross-correlation, which restricts the calculation of the cross-correlation coefficient to only the overlapping regions of Fourier space between any two subtomograms being correlated (Förster et al., 2008). This is done by computing a wedge mask for each of the volumes to be compared, which is a volume which contains zeros in regions of missing information and positive values in other regions. The wedge masks of the subtomograms to be correlated are rotated based on the Euler angles previously found in the subtomogram alignment step, and a compound wedge mask is calculated by multiplying the rotated wedge masks. The subtomograms are then normalised, rotated by their respective Euler angles and optionally multiplied by a region of interest (ROI) mask. The Fourier transform of each rotated subtomogram is calculated and the only pixels in Fourier space within the non-zero parts of the compound wedge mask are used for calculation of constrained cross-correlation. The constrained cross-correlation matrix can then be decomposed by SVD in order to obtain its eigenvectors and eigencoefficients, as described above for the single particle case. However, this method excludes any differences outside the region of Fourier space overlap between the two images from the comparison, and thus does not perform well when the region of overlap is very small (Förster et al., 2008), such as when images taken at high tilt angles have had to be excluded from the data set due to failure of the microscope software to track the region of interest between tilts.

One method that makes use of all of the data while still correcting for the missing wedge is principal component analysis (PCA) of wedge-masked difference maps (WMD), referred to here as WMD PCA (Heumann et al., 2011). In this method, each subtomogram in the data set is rotated by the Euler angles found from alignment against a reference, usually the average of all subtomograms, and multiplied by an optional ROI mask to restrict classification to a certain region in the map. The wedge mask for the corresponding subtomogram is also rotated using the same Euler angles. Both the average structure and the rotated subtomograms are normalised, Fourier transformed and multiplied by the corresponding rotated wedge mask for each respective subtomogram and then inverse-Fourier transformed to produce missing wedge-corrected volumes. A difference map is then obtained by subtraction of the corrected subtomogram from the corrected average structure, and stored as a column in the $m \times n$ matrix D , where m is the number of voxels under the ROI mask for difference map calculation and n is the number of subtomograms. Singular value decomposition of the matrix as follows:

$$D = USV^T$$

yields the matrix U , containing the left singular vectors of D as columns, and the corresponding singular values in the rows of SV^T . The left singular vectors can then be

reshaped to correspond to the original subtomogram pixel coordinates, yielding eigenvolumes that can be used to visually select eigencoefficients (singular values) for clustering by one of the algorithms listed above (Heumann et al., 2011). In Chapter 3, I describe work I have done to implement this classification algorithm in the Briggs group subtomogram averaging pipeline, which I then used to classify different multimeric states from hexamers at the edges of immature Gag lattices.

The structural mechanism of HIV-1 maturation

The work described in this chapter has been published in:

***Mattei, S., *Tan, A.,** Glass, B., Müller, B., Kräusslich, H.-G., & Briggs, J. A. G. (2018). High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proceedings of the National Academy of Sciences*, 201811237. doi:10.1073/pnas.1811237115

*** *Equal contribution***

Some text and figures from the paper above, originally written and produced by myself for the manuscript, have been reproduced in this chapter. See the Publications and Permissions section for more information.

In recent years, much progress has been made in characterising both the immature and mature forms of the hexameric lattice formed by HIV-1 CA. Cryo-electron tomography and subtomogram averaging have given rise to near-atomic resolution structures of the immature CA domain lattice in Gag (Schur et al., 2016), and sub-nanometer structures of the mature CA hexamer and pentamer (Mattei et al., 2016a) from within purified, native viruses. These have enabled the building of atomic models that have revealed the different interfaces involved in stabilising both immature and mature CA lattices in authentic viruses, and provided insights into the mechanism of a class of antiretroviral

drugs known as maturation inhibitors.

However, many aspects of HIV-1 maturation still remain unclear, especially how the transition from the immature form of CA to the mature form occurs. The link between the different proteolytic cleavages that occur in the Gag polyprotein as part of maturation and the structural rearrangement of CA to form a mature capsid is also poorly characterised. Attempts have been made to elucidate these aspects of maturation in the past (de Marco et al., 2010), but the low resolution obtainable by subtomogram averaging at the time due to technological limitations meant that various mechanistic questions remain open.

In this chapter, I describe work that I performed using a panel of mutant Gag constructs, otherwise known as Gag cleavage mutants, in order to determine the structural basis for maturation of an immature CA lattice. I show that formation of a beta-hairpin domain at the N-terminus of CA upon proteolytic cleavage, which was previously thought to be the main determinant of structural maturation, is actually dispensable for maturation in the virus. Instead, the structures I obtained are consistent with unfolding of a six-helix bundle between the C-terminal end of CA and the neighbouring SP1 being the structural switch gating CA maturation.

2.1 Introduction

2.1.1 HIV-1 maturation

As previously explained in Chapter 1, HIV-1 maturation is defined in terms of three distinct sets of processes involved in formation of the mature virus. The first, proteolytic maturation, is defined as the cleavage of the Gag polyprotein at five different sites by the HIV-1 protease, which is a prerequisite for assembly of mature viruses and acquisition of infectivity (Göttlinger et al., 1989; Kaplan et al., 1993; Kohl et al., 1988). The CA domain within Gag that drives immature viral assembly by forming a curved lattice of hexamers is released from the other domains of Gag by proteolytic maturation, allowing it to undergo structural rearrangements.

The rearrangement of the immature hexamer into the mature form of the hexamer, which involves large relative movements of the N-terminal domain (NTD) relative to the C-terminal domain (CTD) and is stabilised by an entirely different set of inter- and intra-hexamer interactions, is called structural maturation (Figure 2.1) (Briggs et al., 2009; Li et al., 2000). This involves loss of a six-helix bundle formed between the end of the CA-CTD and the first part of SP1 (Accola et al., 1998; Schur et al., 2016; Wagner et

al., 2016), which is disordered in the mature hexamer (Mattei et al., 2016a; Zhao et al., 2013). Structural maturation also involves the formation of a beta-hairpin domain by the N-terminal residues of CA upon cleavage of the linker between MA and CA (von Schwedler et al., 1998).

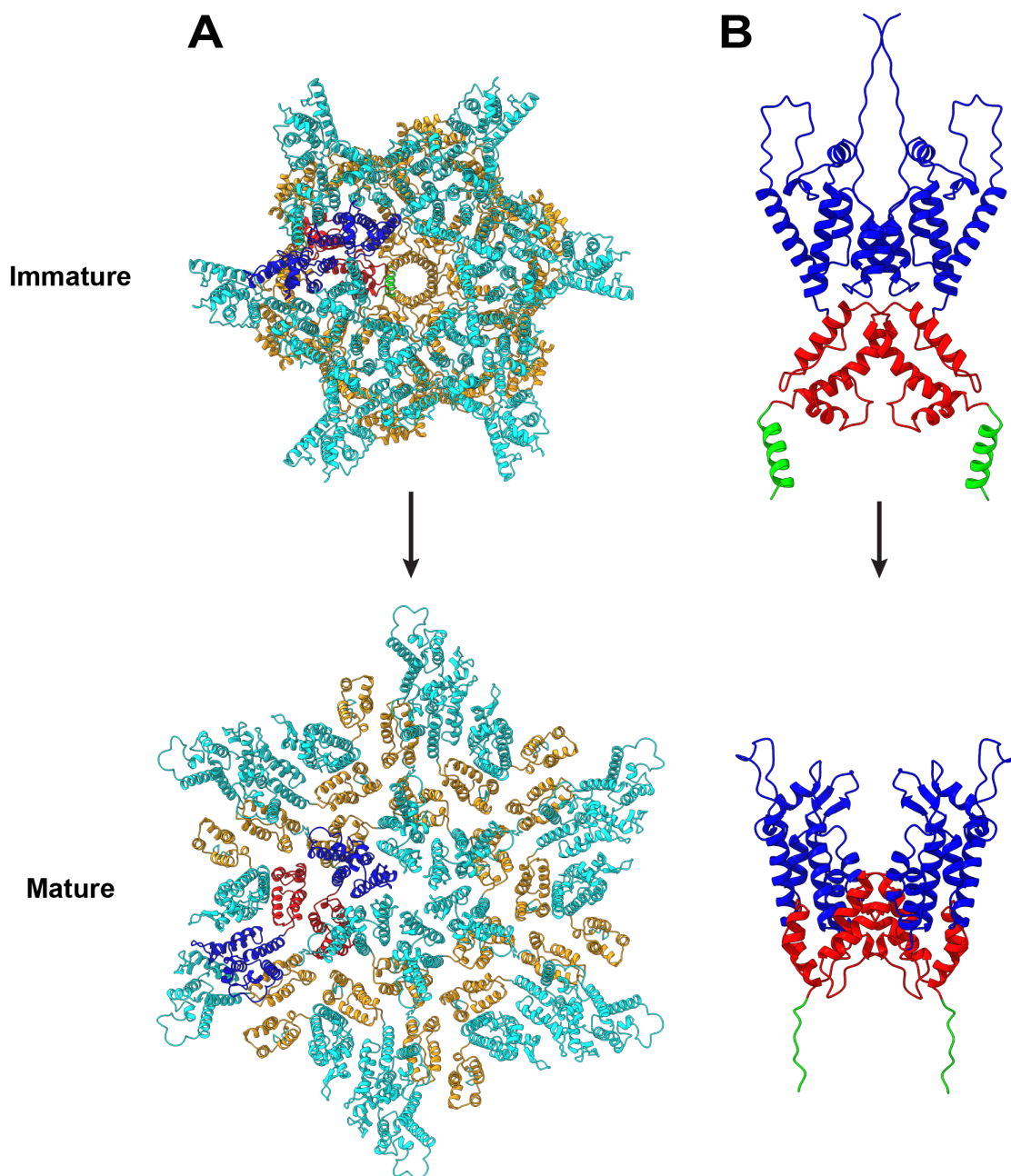


Figure 2.1: Schematic of the structural changes that occur during maturation of HIV-1 CA. **A** Model of the immature (PDB accession code: 5L93) and mature (PDB accession code: 5MCX) hexamer showing the differences in the intra- and inter-hexamer interfaces formed by CA in immature and mature viruses. The CA-NTD is shown in cyan and the CA-CTD in orange. A single dimer in each hexamer is highlighted with the NTD in blue, the CTD in red and the CA-SP1 helix in green. **B** Detailed side view of the immature and mature dimers highlighted in A, showing the large relative movements of the NTD and CTD during maturation.

The third form of maturation is architectural maturation, which describes the morphologi-

cal changes that occur as a consequence of structural maturation. Immature CA hexamers form an incomplete, roughly spherical lattice under the viral envelope, and curvature in this lattice is accommodated by random incorporation of small lattice defects as well as the presence of a large gap in the lattice which is thought to be a budding scar (Carlson et al., 2008). Upon proteolytic cleavage of Gag and subsequent structural maturation, this spherical lattice of CA hexamers rearranges to assemble a closed capsid around the viral ribonucleoprotein (RNP), consisting of a hexagonal lattice of mature CA hexamers that form a fullerene cone by incorporating exactly 12 CA pentamers (Briggs et al., 2003; Mattei et al., 2016a).

In recent years, much progress has been achieved in the structural characterisation of the CA protein in both immature (Schur et al., 2016) and mature (Mattei et al., 2016a) HIV-1, with advances in cryo-EM detector technology, data collection schemes and image processing methods enabling the determination of subnanometer and near-atomic resolution structures of HIV-1 CA assemblies from both immature and mature viruses. These structures have elucidated the very different protein-protein interactions involved in the formation of immature and mature CA lattices and enabled the building of atomic models of the immature and mature hexamer, as well as of the CA pentamer from the mature virus (Mattei et al., 2016a; Schur et al., 2016). However, much is still unknown about the link between proteolytic and structural maturation and the intermediate steps involved in the formation of the mature capsid from an immature Gag lattice are still the subject of much speculation.

Previous studies have attempted to address the question of the relative importance of the proteolytic cleavages upstream and downstream of CA and SP1, but were unable to conclusively deduce a role for each cleavage in modulation of the structural transition from immature to mature CA (de Marco et al., 2010; Wiegers et al., 1998). Both of these regions have been proposed to act as structural switches that regulate maturation (Datta et al., 2011; von Schwedler et al., 1998), but it is unclear whether either, or indeed both of them act in this role or are merely structural changes occurring as a result of maturation. In this chapter, I applied cryo-electron tomography and subtomogram averaging in order to determine high resolution structures of immature and mature CA assemblies in HIV-1 Gag cleavage mutants, viral constructs containing proteolytic cleavage site mutations in different combinations. This has enabled me to ascertain the role of different upstream and downstream cleavages in the initiation of structural maturation, and has revealed the main structural determinant of immature CA domain maturation.

2.2 Aims

I conducted the research in this chapter in order to address the following questions:

1. How does structural maturation depend on proteolytic maturation at the different cleavage sites in Gag, and are specific cleavages absolutely required for maturation to occur? How does this affect architectural maturation?
2. What is the role of the beta-hairpin domain in maturation and does it act as a structural switch as previously suggested?
3. How does proteolytic maturation affect the CA-SP1 six-helix bundle, and can it remain ordered in the mature lattice?
4. What is the relative importance of the beta-hairpin domain and CA-SP1 helical bundle in maturation?

2.3 Materials and Methods

The text in this Materials and Methods section was originally written by myself, except where stated otherwise for individual portions, for publication in the paper Mattei et al. (2018), on which I am co-first author. The text has been reproduced in modified form here from the paper above. See Publications and Permissions section for details.

Viral preparations were made by Bärbel Glass. Cryo-EM grid preparation was done by Simone Mattei. Cryo-ET data acquisition was done jointly by myself and Simone Mattei. All image processing was done by myself, except for structural determination of CA from mature MA-CA and CA-SP1 viruses, which was done by Simone Mattei.

2.3.1 Production and purification of mutant virus particles

This section was originally written by Barbara Müller and modified by me.

“ Particle purification and production: Virus-like particles were produced in HEK293T cells transfected with pCHIV (Lampe et al., 2007) derived plasmids carrying the indicated PR cleavage site mutations (Konvalinka et al., 1995; Müller et al., 2009). For this, cells seeded in 175 cm² flasks were transfected with 70 µg plasmid/flask using a standard CaPO₄ transfection procedure. At 36 h post transfection, tissue culture supernatants were

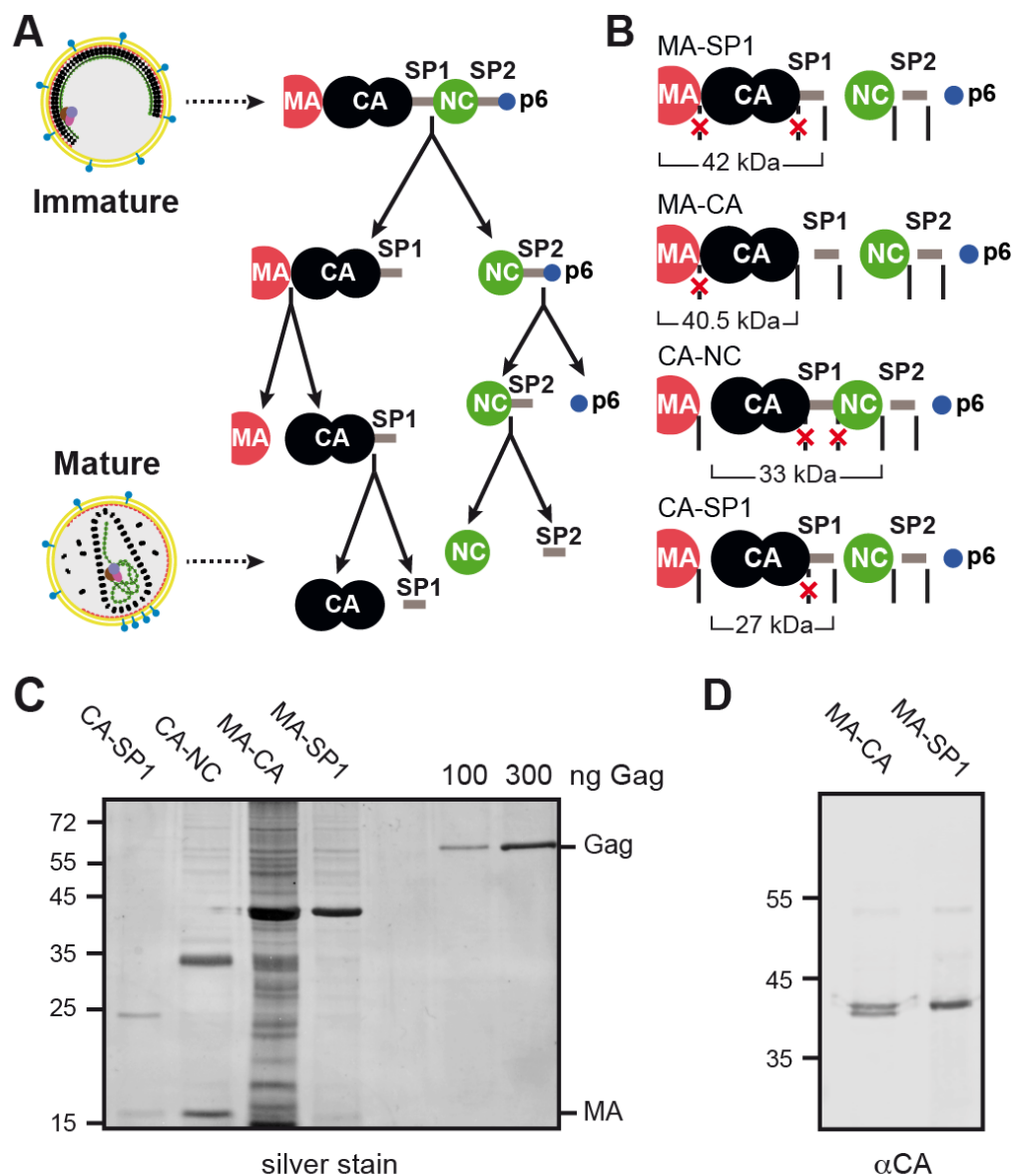


Figure 2.2: Schematics and biochemical characterisation of the cleavage mutant constructs used. **A** Schematic of the proteolytic cleavages involved in HIV-1 maturation, arranged in order of relative rate as determined by Pettit et al. (2005), and the resulting morphological change in the virus from the immature form to the mature form with a closed capsid. **B** Schematic showing the Gag cleavage mutant constructs used with the corresponding molecular weight of the final CA-containing fragment shown below each construct. Cleavage sites at which the protease is able to cleave Gag are indicated by solid black lines, whereas sites which have been mutated so that the protease cannot cleave them are shown using dashed, crossed out lines. **C** SDS-PAGE analysis of the different cleavage mutant constructs used, with numbers to the left of the gel indicating positions of the molecular mass standard bands. Proteins were visualised by silver-staining of the gel, with the two right-most lanes showing the purified, recombinant Gag protein standard used for estimation of relative concentration. **D** Analysis of the cleavage state of the MA-CA and MA-SP1 mutants by quantitative immunoblot, showing that cleavage of the MA-CA mutant occurs at only about 50 % efficiency. Figure originally produced by me for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

The experiments that produced the data for panels C and D were performed by Bärbel Glass at the Department of Infectious Diseases, Universitätsklinikum, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany.

harvested, filtered through 0.45 μm nitrocellulose, and particles were concentrated by ultracentrifugation through a 20 % (w/v) sucrose cushion (2 h at 28 000 r.p.m. in a Beckman SW32 rotor [Beckman Coulter Life Sciences, Indianapolis, Indiana]). Particles were further purified by ultracentrifugation through an iodixanol gradient as described previously (Dettenhofer & Yu, 1999). Particle-containing fractions were diluted with PBS (1:8) and again concentrated by ultracentrifugation (30 min, 44 000 r.p.m., SW60 rotor). Particle pellets were gently resuspended in phosphate buffered saline (PBS), fixed with 1 % paraformaldehyde for 1 h at 0 °C and stored in aliquots at –80 °C.

Immunoblot analysis: Particle lysates were separated by SDS-PAGE (acrylamide : bisacrylamide 200:1) and proteins were transferred to a nitrocellulose membrane by semi-dry blotting. Gag derived proteins were detected by quantitative immunoblot analysis on a LI-COR Odyssey CLx infrared scanner (LI-COR Biotechnology, Lincoln, Nebraska) following the manufacturer’s instructions, using rabbit polyclonal antiserum raised against recombinant HIV-1 CA and IRDye 800CW donkey anti-rabbit IgG secondary antibody (LI-COR Biotechnology). ImageStudio software (LI-COR Biotechnology; version 5.2) was used for generation of images and quantitation of band intensities. ”

Written by Barbara Müller and modified by me.

2.3.2 Cryo-EM grid preparation

10 nm diameter colloidal gold beads suspended in PBS were added as fiducials to purified HIV-1 Gag cleavage mutant viral particles at a 1:1 ratio of gold solution to viral suspension. C-Flat 2/2 3C grids were glow discharged on a Pelco EasiGlow glow discharger with a current of 20 mA for 30 s immediately prior to use. 2.5 μl of each sample was applied to a glow discharged grid, blotted and vitrified in liquid ethane using an FEI Vitrobot Mark II set to a temperature of 15 °C and 100 % humidity.

2.3.3 Cryo-electron tomography

Tilt series were acquired on an FEI Titan Krios 300 keV transmission electron microscope, equipped with a Gatan Quantum 967 LS energy filter operated at a slit width of 20 eV. All tilt series were acquired using a dose-symmetric tilt scheme (Hagen et al., 2017), with a tilt range between –60° and +60° and an angular increment of 3°. Each tilt image was acquired in super-resolution mode on a Gatan Quantum K2 direct electron detector as a movie with the number of frames per tilt indicated in Table 2.1, and a total dose

| Sample | | HIV-1 MA-SP1 | HIV-1 MA-CA | HIV-1 CA-SP1 | HIV-1 CA-NC |
|----------------------|---|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Acquisition settings | Microscope | FEI Titan Krios | FEI Titan Krios | FEI Titan Krios | FEI Titan Krios |
| | Voltage (keV) | 300 | 300 | 300 | 300 |
| | Detector | Gatan Quantum K2 | Gatan Quantum K2 | Gatan Quantum K2 | Gatan Quantum K2 |
| | Energy-filter | Yes | Yes | Yes | Yes |
| | Slit width (eV) | 20 | 20 | 20 | 20 |
| | Super-resolution mode | Yes | Yes | Yes | Yes |
| | Å/pixel | 1.35 | 1.35 | 1.35 | 1.35 |
| | Defocus range (microns) | -1.5 to -5.0 | -1.5 to -5.0 | -2.0 to -4.0 | -1.55 to -5.05 |
| | Defocus step (microns) | 0.25 | 0.25 | 0.25 | 0.25 |
| | Acquisition scheme | -60/60°, 3°, Dose-symmetric | -60/60°, 3°, Dose-symmetric | -60/60°, 3°, Dose-symmetric | -60/60°, 3°, Dose-symmetric |
| | Total Dose (electrons/Å ²) | ~145 | ~150 | ~100 | ~150 |
| | Dose rate (electrons/Å ² /sec) | ~2.3 | ~2.3 | ~2.4 | ~2.3 |
| | Frame number | 10 | 21 | 5 | 20 |
| | Tomogram number | 65 | 70 | 49 | 88 |

Table 2.1: Summary of data acquisition parameters for all HIV-1 Gag cleavage mutants. Table originally made by myself for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

of $\sim 150 \text{ e}^-/\text{\AA}^2$ across the whole tilt series. Frames were aligned on-the-fly in SerialEM (Mastronarde, 2005), and then Fourier cropped to 3708×3708 pixels resulting in a final pixel size of $1.35 \text{ \AA}/\text{px}$ in the unbinned image stacks.

Tilt image stacks were then sorted by tilt angle using IMOD (Kremer et al., 1996) and a dose-dependent exposure filter was applied in MATLAB using the formula described by Grant and Grigorieff (2015). The defocus of each tilt image was estimated from the non-dose-reweighted stacks using CTFFIND4 (Rohou & Grigorieff, 2015), and tomograms were initially reconstructed in IMOD with 2D CTF correction using ctfphaseflip (Xiong et al., 2009). Unbinned tomograms for generation of the final structures were also reconstructed using 3D CTF correction by CTF multiplication of the input stacks with a strip size of 15 nm and astigmatism correction enabled, using novaCTF (Turoňová et al., 2017). These 3D CTF corrected tomograms were used in the final unbinned subtomogram alignment step.

2.3.4 Image processing

| Sample | | Immature HIV-1 MA-SP1 | Immature HIV-1 MA-CA | Immature HIV-1 CA-NC | Mature HIV-1 MA-CA | Mature HIV-1 CA-SP1 | Mature HIV-1 CA-NC |
|--|---------------------------------------|-----------------------|----------------------|----------------------|--------------------|---------------------|--------------------|
| Gag/CA hexamer structure determination | Viruses | 352 | 368 | 338 | 37 | 74 | 29 |
| | Asymmetric units Set A | 265 716 | 231 492 | 159 444 | 10 314 | 23 988 | 5 844 |
| | Asymmetric units Set B | 265 728 | 231 450 | 159 174 | 10 320 | 23 994 | 5 940 |
| | Final resolution (0.143 FSC) in Å | 4.0 | 3.7 | 4.5 | 8.3 | 7.9 | 9.7 |
| | EMDB accession code | EMD-0164 | EMD-0165 | EMD-0166 | EMD-0167 | EMD-0169 | EMD-0168 |
| CA pentamer structure determination | Asymmetric units | – | – | – | – | 55 | 70 |
| | Final cross-resolution (0.5 FSC) in Å | – | – | – | – | 24 | 22 |
| | EMDB accession code | – | – | – | – | EMD-0171 | EMD-0170 |

Table 2.2: Summary of image processing parameters for all HIV-1 Gag cleavage mutants. Table originally made by myself for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

Immature viruses: Centres of immature viruses in the MA-SP1, MA-CA and CA-NC datasets were marked manually in $8\times$ binned tomograms using a custom plugin for UCSF Chimera (Pettersen et al., 2004), and an oversampled, spherical grid of initial subtomogram extraction points was defined for each virus in MATLAB using these centres and the radius of each virus. These extraction points were used to perform independent subtomogram averaging and alignment for each data set using a set of in-house MATLAB scripts based the TOM (Nickell et al., 2005), AV3 (Förster et al., 2005) and Dynamo (Castaño-Díez et al., 2012) packages, as previously described for wild-type immature HIV-1 particles by (Schur et al., 2016).

One tomogram acquired at high defocus was chosen from each data set for use in constructing an initial alignment reference. This was done for each data set by extracting subtomograms from the $8\times$ binned tomogram with a box size of 389 Å^3 , averaging them and performing iterative reference-free alignment, using cone search parameters of $6\times 4^\circ$ and an in-plane search of $8\times 5^\circ$. 5 iterations with a low pass filter of 29.9 Å were performed for the MA-SP1 data set, and 7 iterations with a low pass filter of 32.4 Å were run for the MA-CA and CA-NC data sets. The position of the average within the box was then shifted to centre the C6 symmetry axis and 2 further iterations were run with the same parameters but with C6 symmetry applied to the reference. A minimum dis-

tance threshold of 32.4 Å (4 px) was then used for removal of duplicate subtomogram positions from the data set, and positions with a low cross-correlation coefficient (CCC) score were also removed to exclude subtomograms that did not align to any Gag protein layer. One final alignment iteration was run on the MA-SP1 data set with a $4 \times 1^\circ$ cone search, $5 \times 1^\circ$ in-plane search and C6 symmetry applied, whereas for the MA-CA and CA-NC data sets, three further iterations were run with the same parameters as before and a C6-symmetrised reference. The average from the final iteration was used as the initial reference for subtomogram alignment of the complete set of tomograms from each respective data set.

For the full data sets, 2D CTF corrected tomograms were used for all alignments unless otherwise stated. Subtomograms were extracted from the full set of extraction points defined for each immature cleavage mutant data set as before, and aligned against the respective reference for 2 iterations with a $7 \times 6^\circ$ angular search for all Euler angles, a 32.4 Å low pass filter and with C6 symmetry imposed. Cleaning of the aligned positions according to pairwise distance and CCC were then performed as described above, using a distance threshold of 8 nm. The remaining positions were then successively aligned using 4× binned, 2× binned and then unbinned subtomograms with the same steps and parameters described by (Schur et al., 2016). The data sets were split into equally-sized odd and even halves after alignment with 4× binned data in order to allow for resolution determination using the gold-standard FSC method.

After the second alignment iteration using unbinned, 2D CTF corrected subtomograms, the centre-of-mass for defocus determination was calculated from the aligned subtomogram positions, for use in 3D CTF estimation and correction. Unbinned tomograms were then reconstructed with 3D CTF correction using these centre-of-mass values as described above, and subtomograms were re-extracted with a box size of 259 Å³ from the 3D CTF corrected tomograms. These subtomograms were averaged to produce new references for the odd and even half data sets, which were then used for one further alignment iteration using a $2 \times 1^\circ$ angular search for all Euler angles, a band pass filter between 8.1 Å and 25.9 Å and C6 symmetry imposed. Image processing statistics are summarised in Table 2.2.

Mature viruses: Complete and incomplete mature cores in the CA-NC data set were segmented in Amira (FEI, versions 5.5 and 6.1.1) by tracing their approximate outlines on XY slices of 8× binned tomograms, and interpolating the traces in Amira to produce binary masks defining the approximate core volume. These volumes were filtered with a circular Gaussian kernel of 20 pixels in MATLAB using the `tom_filter` function from the TOM package, and isosurfaces were defined on the surface of the filtered volumes. The isosurfaces were then used to define a grid of extraction points with a 2 pixel spacing

at an intensity threshold of 0.3, and Euler angles were assigned based on the normal to the isosurface at each extraction point, with random in-plane angles assigned in the range of 0° to 360° .

Initial extraction points for CA-SP1 and MA-CA viruses containing mature lattices were defined in Amira (version 4.1.2) by estimating the curvature of the lattice using a sphere, and marking the centre of this sphere with a point that was then used with this radius to generate extraction points with 2.16 nm spacing as described for the immature viruses. Initial alignment references for the mature CA-SP1 and MA-CA lattices were then constructed from a single $8 \times$ binned tomogram per data set, whereas a single viral core was used for initial mature CA-NC reference generation. Subtomograms were extracted with a box size of 691 \AA^3 and averaged, and then alignment was performed against the reference for the respective data set using a cone search of $2 \times 10^\circ$, $18 \times 10^\circ$ in-plane search, a 49.4 \AA low pass filter and no imposed symmetry until convergence on a hexagonal lattice was observed. Two more iterations of alignment were then run, manually excluding points with low CCC from the average but retaining them for iterative alignment. The C6 symmetry axis of each reference was then centred in the box and two further alignment iterations were run with a $5 \times 4^\circ$ cone search, $5 \times 6^\circ$ in-plane search, low pass filter of 49.4 \AA and C6 symmetry imposed during alignment. The resulting averages from the final iteration were used as the final references for each respective data set.

For alignment of the full data sets of CA-SP1 and MA-CA mature hexamers, subtomograms were extracted from $8 \times$ binned tomograms using the extraction points defined above, and then aligned for two iterations against the respective references previously generated, using a $10 \times 4^\circ$ cone search, $10 \times 3^\circ$ in-plane search, low pass filter of 49.4 \AA and a C6-symmetrised reference. Duplicate and misaligned positions were then removed as described for the immature data sets, using a distance threshold of 5.4 nm, and in the case of the CA-SP1 data set, 2 further iterations of alignment were then run using the same parameters. The mature CA-NC data set was processed differently due to the irregular shape of the mature cores present in this data set, necessitating additional steps to correct the segmentation. All alignments of the CA-NC data set were performed on C6-symmetrised references. Subtomograms were first extracted from $8 \times$ binned tomograms at the coordinates defined from segmentation of mature lattices in the CA-NC data set, using the same box size as used for the mature MA-CA and CA-SP1 data sets. These subtomograms were aligned along the Z-axis only against the initial reference generated earlier, with a $4 \times 10^\circ$ cone search, $3 \times 10^\circ$ in-plane search and 49.4 \AA low pass filter. Removal of subtomograms with low CCC was done as for the other two mature core data sets and the remaining points were converted into a marker set in UCSF Chimera, which was overlaid on the corresponding tomogram for filling of gaps with the Volume Tracer tool where necessary. The corrected marker sets were then used to generate a new isosurface

for extraction point definition in MATLAB as before, and using these revised positions, subtomograms were extracted once again from $8\times$ binned tomograms for alignment as before. Exclusion of duplicate and misaligned positions based on CCC was performed as before, and 2 further iterations of alignment were run with a $4 \times 5^\circ$ cone search and $5 \times 6^\circ$ in-plane search.

All of the mature data sets were then processed independently as follows after alignment with $8\times$ binned subtomograms. Subtomograms were extracted from $4\times$ binned tomograms using a box size of 346 \AA^3 and averaged, and then the subtomograms were aligned against the respective average. Three iterations of alignment were used, with a $4 \times 6^\circ$ search range for each Euler angle and a 34.6 \AA low pass filter. The aligned positions were then scaled for use with $2\times$ binned data, from which subtomograms were re-extracted using a box size of 346 \AA^3 . Subtomograms with grey values beyond ± 1 standard deviation from the mean were excluded in order to remove subtomograms that may have contained projection artefacts from gold beads, and the remaining subtomograms were split into odd and even half data sets for independent processing. The subtomograms in each half data set were averaged to produce odd and even alignment references, which were then used for two iterations of alignment of the $2\times$ binned data with a $3 \times 2^\circ$ search for all Euler angles and a low pass filter of 24.7 \AA . Subtomogram positions were again scaled and subvolumes were then extracted with a box size of 346 \AA^3 from unbinned tomograms, followed by averaging to produce new references for the odd and even half data sets.

One iteration of alignment of the unbinned subtomograms was then run with a $3 \times 1^\circ$ angular search for all Euler angles and a low pass filter of 15.3 \AA . This was followed by a second iteration of alignment with a $2 \times 1^\circ$ angular search for all Euler angles and band pass filter between 13 \AA and 20 \AA . Subtomograms were re-extracted from 3D CTF corrected, unbinned tomograms at the positions obtained from the last iteration of alignment, and this iteration of alignment was repeated once on the 3D CTF corrected data in order to generate the final structure for each data set.

2.3.5 Visualisation of structures and maps

The analysis of immature Gag structures shown in Figure 2.5 was performed in UCSF Chimera by rigid body fitting of the central protomer of an atomic model of the immature CA domain (PDB accession code: 5L93) into the density maps. The equivalent analysis of the mature structures in Figure 2.7 was performed in the same way using a crystal structure of mature CA (PDB accession code: 4XFX) for fitting. A crystal structure of mature CA (PDB accession code: 5HGK) and an NMR solution structure of an MA-containing N-terminal fragment of immature Gag (PDB accession code: 1L6N) were also

fit into the immature structures in order to assess the extra N-terminal density in the CA-NC map to determine its compatibility with a beta-hairpin structure.

Isosurface renders of all structures, as well as orthoslices through $4\times$ binned data in the case of the pentamer structures, were generated using UCSF Chimera. All orthoslices of viruses shown in Figures 2.3 and 2.4 were generated from CTF-corrected, $4\times$ binned tomograms which were filtered with a Gaussian kernel of 5 binned pixels in MATLAB to aid visualisation, and were rendered at a thickness of 5.4 \AA . Pentamer structures obtained from the mature cleavage mutant viruses were compared to a pentamer structure from the wild-type virus, which was previously also generated using $4\times$ binned data by Mattei et al. (2016a), in order to facilitate a more direct comparison of the maps. PyMOL (“The PyMOL Molecular Graphics System, Version 1.8.6.0 Schrödinger, LLC.”, 2017) was used to assess the fit of a single Gag subunit (PDB accession code: 5L93) into the high resolution maps of the immature hexamer (Figure 2.6).

The mature and immature CA hexamer references used were different enough that no subtomogram positions were observed to align to both references, and in the case of viruses that contained both immature and mature lattices, the following lattice map analysis was performed separately on subtomograms aligned to the mature and immature references, respectively. Aligned subtomogram positions and orientations were used to generate ‘lattice maps’ using a custom plugin for UCSF Chimera, as described by Qu et al. (2018), in order to perform the analysis shown in Figure 2.4 B-C. This was done by rotating a geometric object using the Euler angles of each subtomogram as found during alignment and placing it at each of the corresponding subtomogram extraction coordinates within the tomogram, in order to visualise the arrangement of hexamers within the mature and immature CA lattices. The coordinates used for visualisation were generated during the final iteration of alignment on $8\times$ binned data, following which misaligned and duplicate subtomogram positions were removed based on CCC and pairwise distance. The CCC criterion for subtomogram removal was set on a per-tomogram basis and chosen such that obviously misaligned subtomograms, defined as those with translations or rotations that caused them to no longer correspond to the CA protein layer, were excluded while ensuring that lattice completeness was not compromised by removal of false negatives. Further outlier positions that could not be removed using CCC were then manually removed in UCSF Chimera, with the criterion for removal being misalignment, as defined above, or the positions not being continuous with any region of the hexameric lattice.

2.4 Results

2.4.1 Large data sets reveal novel minority phenotypes

The large amount of data acquired here (Tables 2.1 and 2.2), which is of significantly better quality than previously obtainable by de Marco et al. (2010) due to improvements in detector technology and data processing, has allowed a more comprehensive analysis of both the morphology and structure of these mutant viral constructs. This has revealed the existence of previously unknown minority phenotypes (Figure 2.3).

de Marco et al. (2010) previously concluded that cleavage on both ends of the CA-SP1 module is required for maturation to occur, as only immature viruses were reported in their MA-SP1, MA-CA and CA-p6 (identical to CA-NC except for the p6 domain) data sets. Contrary to these previous results, here I show that structural maturation of CA can occur when cleavage is only possible on one end of CA. This is shown by the presence of MA-CA and CA-NC viruses that possess mature CA lattices (Figure 2.3, panels ix-x, xiv-xv).

However, maturation in these cleavage mutant constructs occurs at very low (<10 %) frequency, compared to a majority of mature viruses when both ends of CA-SP1 can be cleaved, as is the case in CA-SP1 (Figure 2.3, panels xvii-xx). Notably, some of the MA-CA and CA-NC viruses have both immature and mature lattices in close juxtaposition (Figure 2.3, panels ix and xiv) with some density in between, although attempts to perform subtomogram alignment on this density between the two lattice types did not reveal any ordered structure there.

2.4.2 Morphological analysis of HIV-1 Gag cleavage mutants

Immature viruses

As described in Materials and Methods, I applied reference-free subtomogram averaging in order to characterise both the structure and morphology of the immature and mature lattices seen in the cleavage mutant data sets. Analysis of the lattice maps, which are aligned subtomogram positions and orientations plotted as geometric objects within the original tomograms, show that the majority of immature viruses in the MA-SP1, MA-CA and CA-NC data sets exhibited Gag lattice architecture typical of the wild-type virus, namely a lattice of hexamers in a truncated sphere containing small defects as well as a large gap at

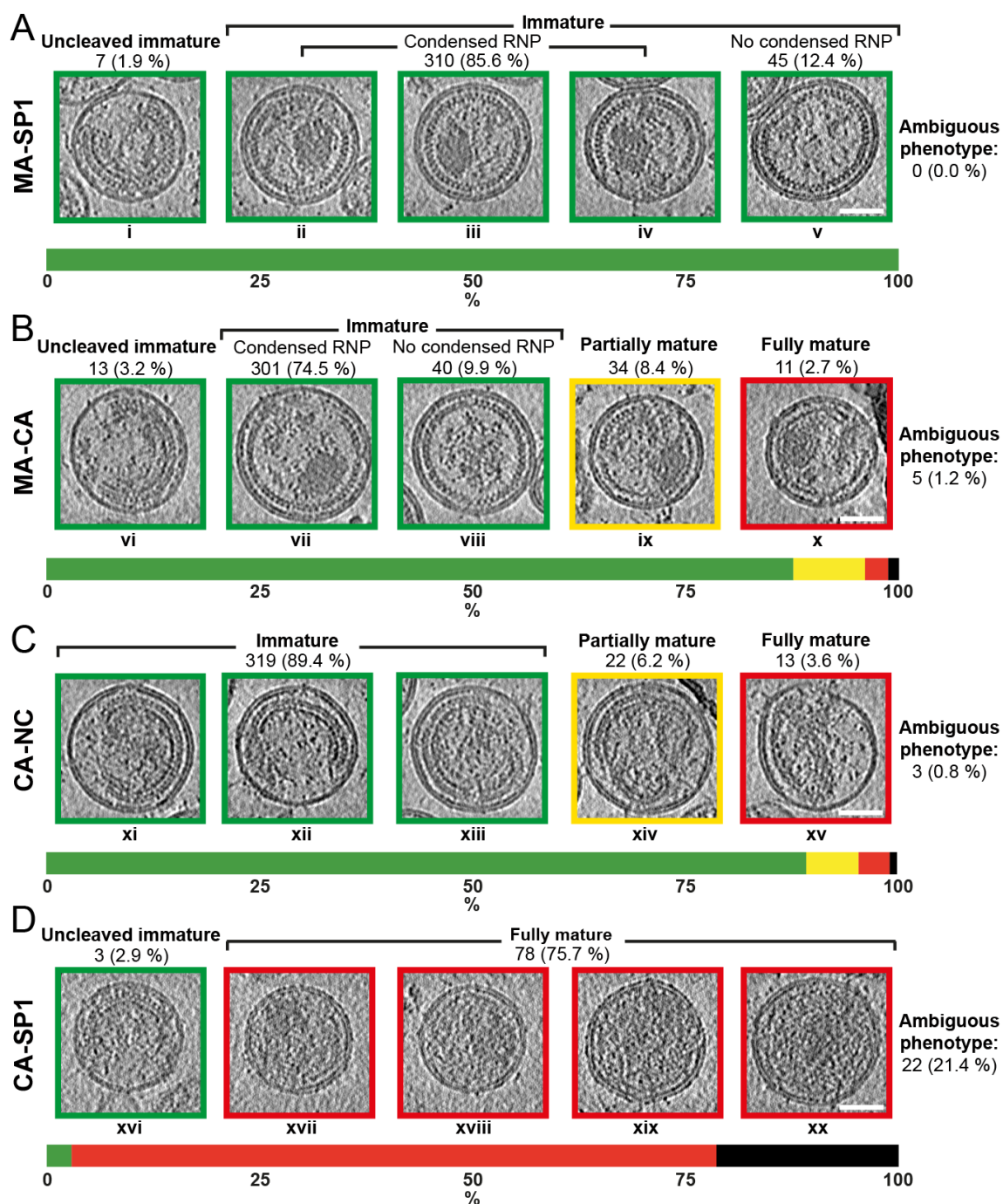


Figure 2.3: Distribution of phenotypes observed in each of the Gag cleavage mutant data sets. Scale bars = 50 nm. Computational slices of 5.4 Å thickness through Gaussian-filtered, 4× binned tomograms are shown, illustrating the different viral phenotypes exhibited by the different Gag cleavage mutant viral constructs. Green boxes denote immature lattice morphology, red boxes indicate that the CA lattice is mature and yellow boxes mean that both immature and mature lattices are present in the same virus. Colour bars under the images for each data set denote the percentage of the viral population accounted for by each of these phenotypes. Figure originally produced by me for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

the budding site (Figure 2.4, panels i-iii, iv-vi and xvi to xviii). As expected, the density corresponding to the MA-SP1 and MA-CA immature lattices appeared thinner due to the lack of the NC-RNA layer, but also contained a dark, luminal density corresponding to the condensed RNP (Figure 2.4, panels i-iii and iv-vi).

However, a subset of the MA-CA and MA-SP1 viral particles contained a much more complete immature lattice which did not have a large gap and instead covered almost the entire inner surface of the virus (Figure 2.3, panels v and viii). A condensed RNP density was not detectable in many of these viruses. This phenotype has previously been reported in viruses produced by expression of Gag and Gag-Pol from an adenoviral vector, and was attributed to premature SP1-NC cleavage during budding that resulted in loss of the p6 domain (Carlson et al., 2010). Budding in such a case would need to occur via an ESCRT-independent pathway, which would allow the immature lattice to continue growing until it is almost complete. The observations here from the MA-SP1 and MA-CA mutants are consistent with the conclusion of Carlson et al. (2010) that such a phenotype is due to premature proteolytic processing between the SP1 and NC domains.

Mature viruses

A surprising result from the analysis of these cleavage mutant viral constructs is that a minority of MA-CA and CA-NC viruses contained mature CA lattices, including viruses that contained both immature and mature lattices in close apposition. Both of these phenotypes were not previously detected when these constructs were first structurally analysed by de Marco et al. (2010), due to the much smaller size and poorer image quality of those data sets. The majority of CA-SP1 cleavage mutant viruses were mature (Figure 2.3, panels xvii-xx), which agrees with a previous study carried out by de Marco et al. (2010) on this mutant construct. In the case of CA-NC and CA-SP1, a small number of pentamers were also detected by subtomogram alignment, but were not seen in MA-CA, possibly due to the different curvature of the MA-CA lattice enforced by the uncleaved linker between MA and CA (Figure 2.4, panel xxiv).

Analysis of the lattice maps obtained by subtomogram averaging of these lattices shows that architectural maturation in all of these mutants was defective, as most of the mature CA lattices did not successfully form closed, conical cores (Figure 2.4, panels xiv-xv, xxiii-xxv and xxvi-xxvii), instead taking the form of irregular polyhedra that were in many cases open. This was particularly apparent in the case of MA-CA, in which the linker between the MA and CA domains remained uncleaved, thus keeping the CA layer tethered to the membrane. Mature MA-CA viruses exhibited a lattice morphology that is very similar to the immature lattice, that is, hexagonal tiling to form a truncated sphere with

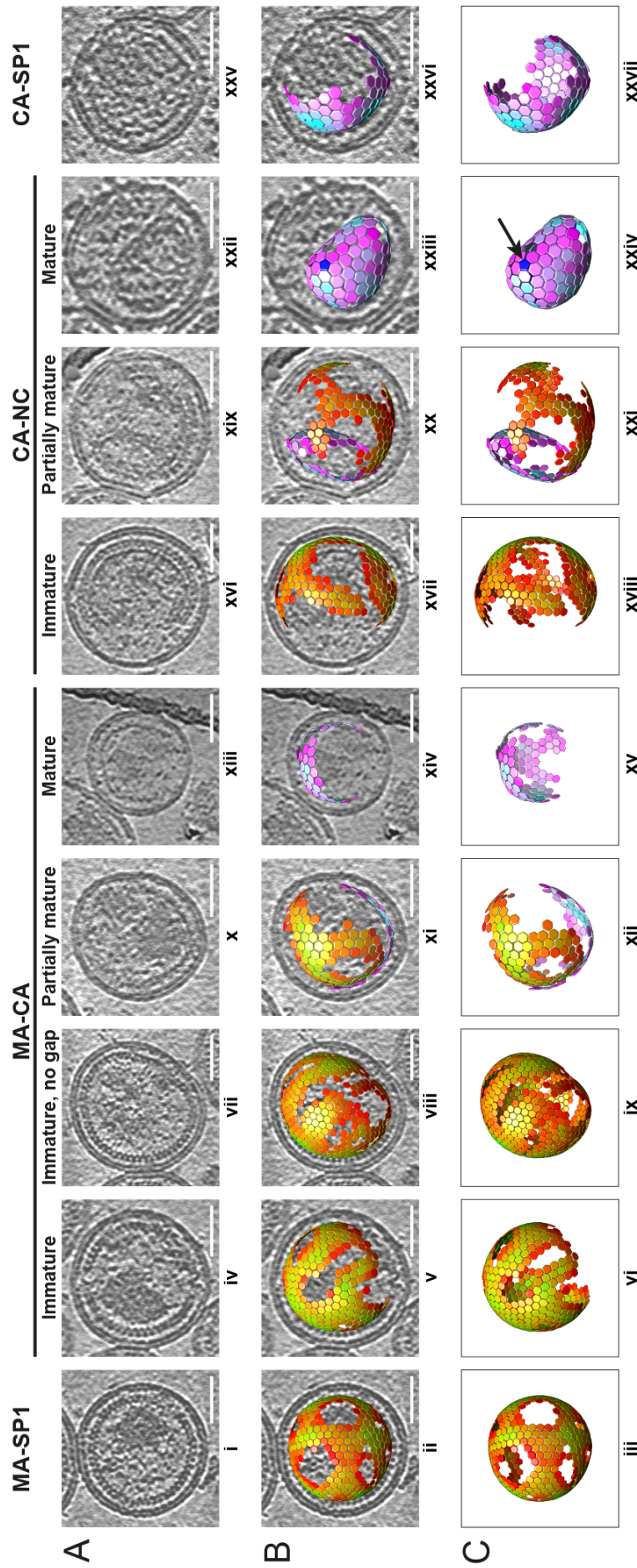


Figure 2.4: Lattice map representation of CA architecture in each cleavage mutant. Scale bars = 50 nm. **A** Computational slices of 5.4 Å thickness through Gaussian-filtered, 4x binned tomograms showing representative viruses of each observed phenotype in all of the Gag cleavage mutant data sets. **B** As in A, with lattice maps overlaid to show the aligned positions and orientations of subtomograms used for structural determination. Colour ranges, from low to high cross-correlation coefficient (CCC); red-green = immature hexamers, magenta-cyan = mature hexamers, dark blue = mature pentamers. **C** As in B, showing only the lattice maps. Figure originally produced by me for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

both small defects and a large gap. Subtomogram alignment clearly revealed these lattices to be structurally mature despite this immature-like architecture (Figure 2.4, panels xi-xii and xiv-xv). This highlights the distinction between the different types of maturation. Although cleavage on either end of CA is sufficient for some structural maturation to occur, architectural maturation is defective in either case. Even in the case of CA-SP1, where cleavage occurs on both ends of the CA-SP1 module, the majority of viruses did not form closed cores, instead adopting a more irregular, open structure (Figure 2.4, panels xxvi-xxvii), which points to the importance of kinetic regulation of structural maturation in achieving correct architectural maturation.

2.4.3 Structures of immature CA in cleavage mutants

Using subtomogram averaging, I determined the structure of the CA hexamer in the immature Gag lattice of the MA-SP1, MA-CA and CA-NC cleavage mutant viruses to near-atomic resolution (Figure 2.5). These structures show that the overall quaternary structure of the immature CA hexamer in each of these cleavage mutants is essentially identical to that in the wild-type virus (Figure 2.5A). The MA-SP1 and MA-CA cleavage mutants, in particular, did not exhibit any structural changes in the six-helix bundle formed between the carboxy-terminal end of CA and the first part of SP1, with the length of the ordered part of SP1 remaining exactly the same as in the wild-type virus regardless of proteolytic cleavage state (Figure 2.5). This indicates that cleavage between the SP1 and NC domains does not give rise to any detectable structural changes in the immature CA hexamer, and also that the six-helix bundle is ordered in each of the proteolytic cleavage states still compatible with immature lattice formation.

It is interesting to note that this helical bundle is of the same length in both the MA-SP1 and MA-CA mutants, which only differ in whether the cleavage site between CA and SP1 within the bundle has been mutated, and therefore no structural change can be attributed to removal of the SP1 domain. Immunoblot analysis of MA-CA virus preparations consistently show that about 50 % of the cleavage sites between CA and SP1 are uncleaved (Figure 2.2D), which may explain this phenotype.

However, the CA-NC cleavage mutant exhibited one notable structural difference compared to the wild-type immature hexamer, in the form of an extra density upstream of helix 1 in its N-terminal domain (Figure 2.5B and C). A rigid-body fit of a crystal structure of helix 1 and the adjacent N-terminal beta hairpin domain (PDB accession code: 5HGK) into the immature CA-NC density map shows that this extra density is consistent with the existence of a beta-hairpin in the immature CA-NC hexamer. This is a surprising result, as the beta-hairpin is characteristic of the mature CA hexamer and has been suggested

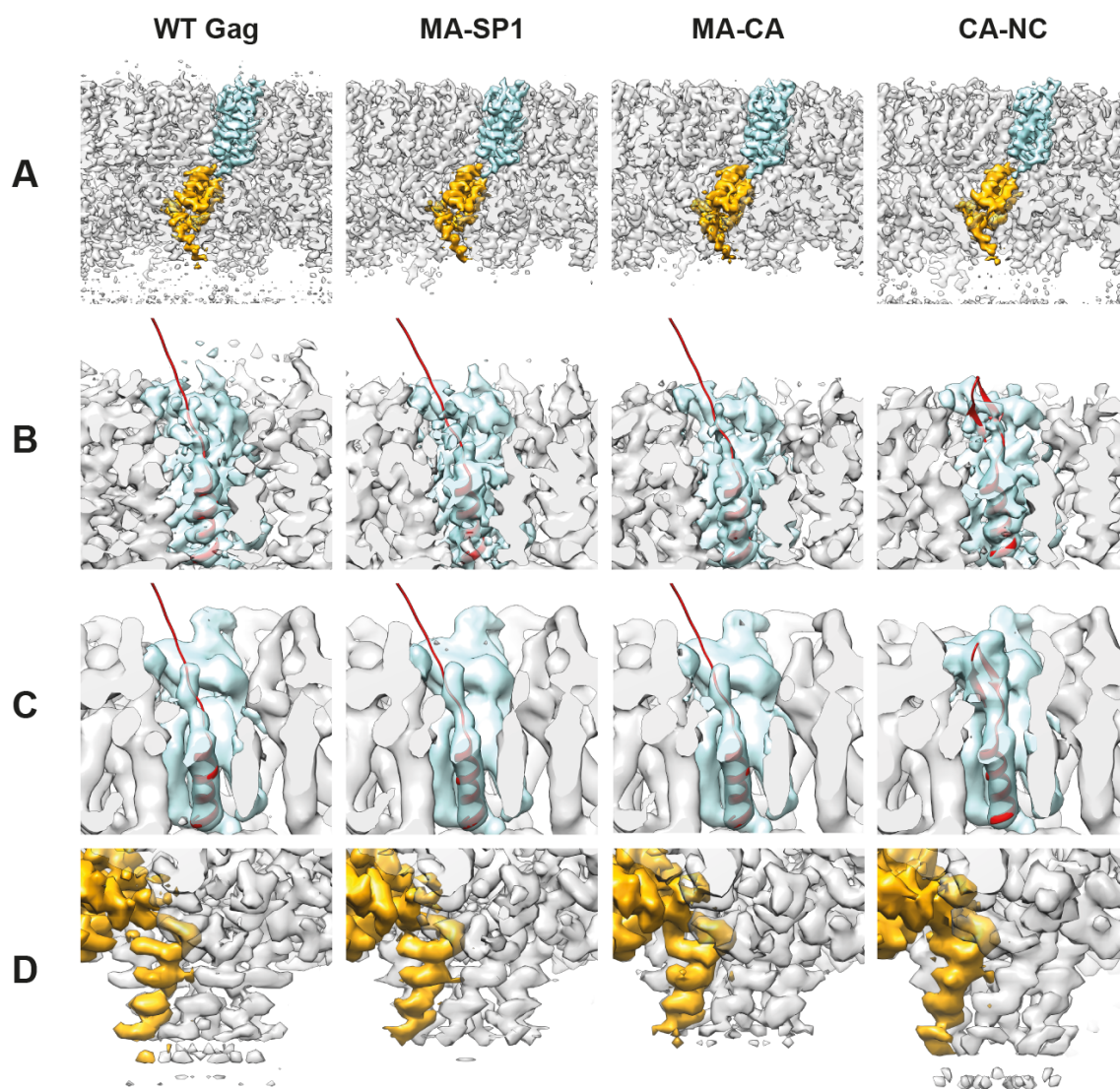


Figure 2.5: Immature hexamer structures from Gag cleavage mutant viruses. The first column depicts the structure of the CA hexamer in the immature Gag lattice of the wild-type (WT) virus, determined by Schur et al. (2016) to 4.2 Å resolution (EMDB accession number: EMD-4017). The corresponding regions of the immature hexamer maps in MA-SP1, MA-CA and CA-NC, resolved to 4.0 Å, 3.7 Å and 4.5 Å respectively, are shown in the other columns. **A** Side view of a slice through an isosurface render of the immature hexamer in each of the maps, with a single CA protomer highlighted for emphasis (cyan = NTD, orange = CTD). **B** Close up view of the density in the CA-NTD corresponding to helix 1 (cyan) and linker upstream of this helix to the MA domain. For the WT, MA-SP1 and MA-CA maps, an NMR solution structure of this helix and linker (PDB accession code: 16LN) are fit into the map, showing no ordered density upstream of helix 1 that can be attributed to the linker. The CA-NC map has an atomic model of helix 1 with a beta-hairpin, derived from a crystal structure of mature CA (PDB accession code: 5HGK) fit instead. The CA-NC map has a clear extra density into which a model of the beta-hairpin easily fits as a rigid body. **C** As in B, but with all maps low pass filtered to 6.0 Å, which makes the extra density at the N-terminus of CA in the CA-NC map more apparent. **D** Side view of the 6-helix bundle between the C-terminus of CA and the first part of SP1. Regardless of proteolytic cleavage state, the density in this region is identical to that in WT immature Gag. Figure originally produced by me for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

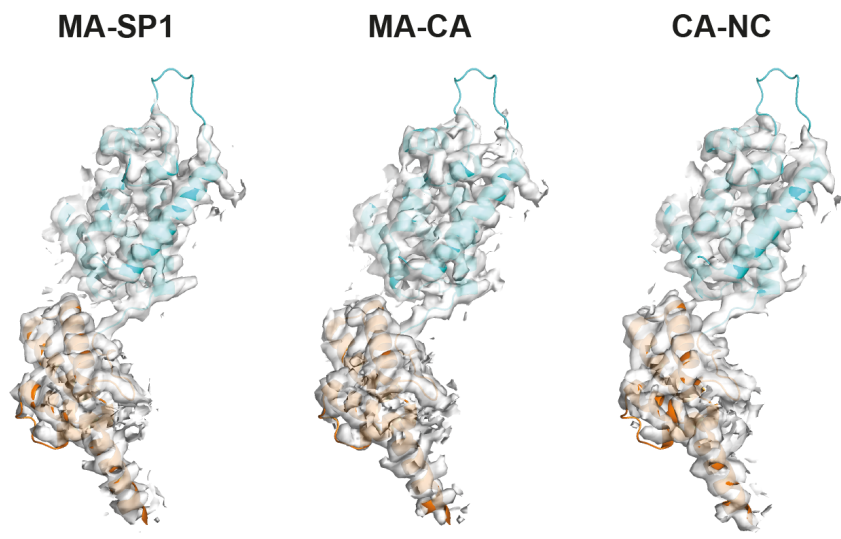


Figure 2.6: Fit of the wild-type CA atomic model from immature Gag into immature cleavage mutant hexamer maps. A single chain from the atomic model from the wild-type immature hexamer (PDB accession code: 5L93) was fit into the immature hexamer density maps from the MA-SP1, MA-CA and CA-NC cleavage mutants, showing that the overall conformation of immature CA in each cleavage mutant is the same as in the wild-type virus.

to destabilise the immature lattice and therefore act as a structural switch for maturation (von Schwedler et al., 1998). This result shows that the beta-hairpin is clearly compatible with the immature form of the CA hexamer, and cannot be the main determinant of maturation.

2.4.4 Structures of mature CA in cleavage mutants

This section describes work that was jointly performed by myself and co-first author, Simone Mattei, on the published paper (Mattei et al., 2018). I determined the structure of the mature hexamer and pentamer in the CA-NC cleavage mutant, and Simone Mattei determined the hexamer and pentamer structures in mature MA-CA and CA-SP1 viruses. Fitting of atomic models into each of these maps and subsequent interpretation were carried out by me.

Mature CA hexamer structure determination

As with the immature viruses, structural determination of the mature lattices found in the MA-CA, CA-NC and CA-SP1 cleavage mutants was carried out by reference-free subtomogram averaging, although the resolution achieved was significantly lower than with the immature cleavage mutants due to the much smaller amount of data available

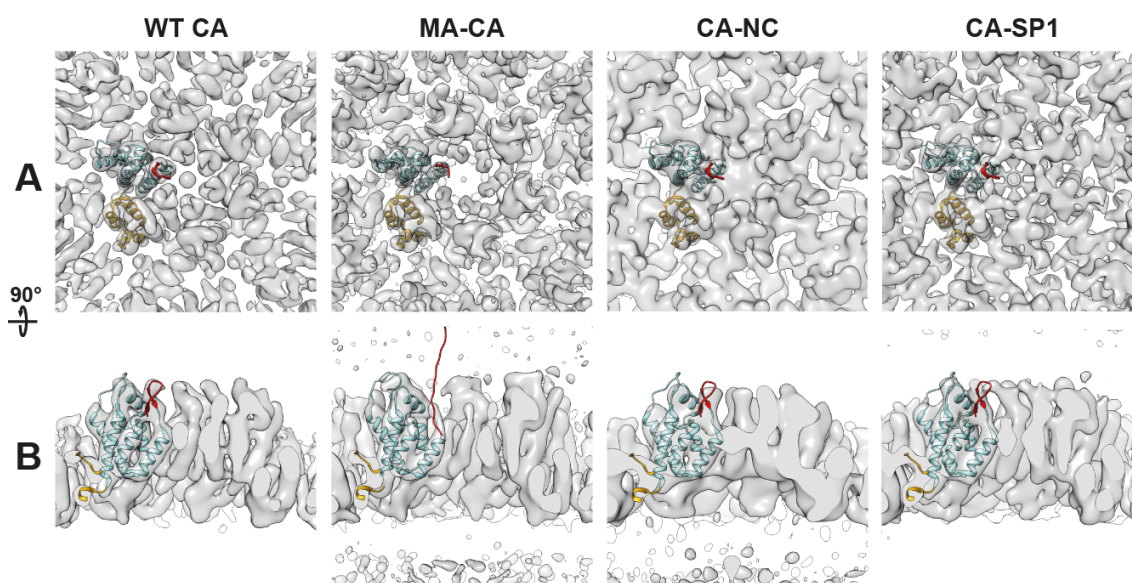


Figure 2.7: Mature hexamer structures from Gag cleavage mutant viruses. The wild-type mature CA hexamer structure, determined from authentic viral particles by Mattei et al. (2016a) to 6.8 Å resolution (EMDB accession number: EMD-3465), is shown for comparison, with the MA-CA, CA-NC and CA-SP1 mature hexamer structures resolved to 8.6 Å, 9.7 Å and 7.9 Å respectively. **A** Isosurface view of the hexamer in the wild-type virus and the cleavage mutants as seen from the direction of the NTD towards the CTD. A backbone atomic model derived from the WT hexamer map by Mattei et al. (2016a) is shown (PDB accession code: 5MCX), fit into the structure as a rigid body. Colour key: cyan = NTD, orange = CTD, red = beta-hairpin. Instead of a beta-hairpin model, the MA-CA map shows an NMR solution structure of the linker between MA and CA (PDB accession code: 16LN). **B** Side view through the same hexamer maps as in A, clearly showing that the MA-CA mature hexamer map lacks the beta-hairpin density seen in the other two mutants and the WT virus. All of the maps do not have ordered density corresponding to the residues that form the 6-helix bundle between CA and SP1 in the immature hexamer, regardless of proteolytic cleavage state. Figure originally produced by me for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

(Table 2.2). However, the resolution to which each map was obtained was still in the sub-nanometer range and sufficient to resolve secondary and tertiary structure in each of the hexamer maps (Figure 2.7). These maps showed that the overall arrangement of CA in the hexamer was similar to that previously reported by Mattei et al. (2016a) in the wild-type virus, regardless of the cleavage state of the CA-containing fragment. In particular, the CA-NC cleavage mutant did not contain any additional structured regions beyond helix 11 of CA, showing that the CA-SP1 region, which forms a helical bundle in the immature structure, is unstructured in the mature hexamer even when no cleavage is possible between CA, SP1 and NC (Figure 2.7B).

The beta-hairpin density was also resolved in the CA-NC and CA-SP1 hexamer maps, although the conformation of this domain could not be reliably ascertained due to the resolution of the structures. However, one striking difference was seen in the mature hexamer present in the MA-CA cleavage mutant. Unlike the CA-NC and CA-SP1 cleavage mutants, in which a beta-hairpin density was clearly resolved upstream of helix 1, there was no corresponding density for the beta-hairpin in the mature MA-CA hexamer (Figure

2.7B). In place of the beta-hairpin density, a much weaker density was resolved that corresponds to the unstructured linker between MA and CA, which cannot be cleaved in the MA-CA mutant. This is direct evidence that the beta-hairpin is dispensable for the formation of a mature hexamer that is otherwise structurally identical to that in the wild-type virus.

Mature CA pentamer structure determination

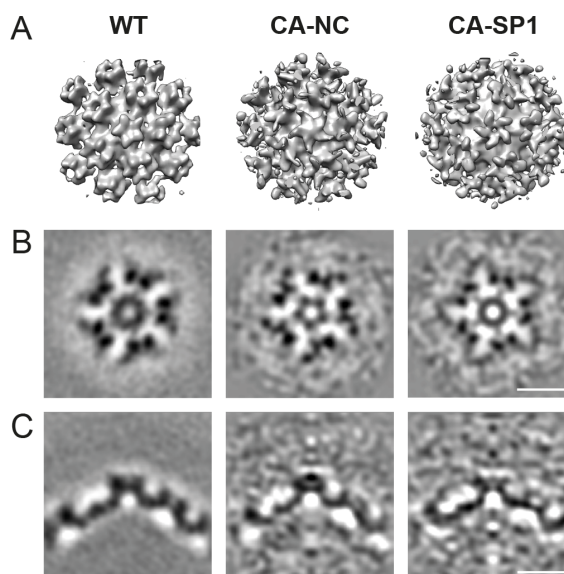


Figure 2.8: Mature pentamer structures from Gag cleavage mutant viruses. The structures are represented as 5.4 Å thick computational slices through the maps with overlaid 10 nm scale bars. A structure of the wild-type viral pentamer (EMDB accession number: EMD-3466) is shown adjacent to the 22 Å CA-NC and 24 Å CA-SP1 pentamer maps for comparison. **A** Isosurface render of the CA pentamer from wild-type, CA-NC and CA-SP1 viruses. **B** Corresponding view as a computational slice through the centre of the map with a thickness of 5.4 Å. To the determined resolution, the pentamer in both of these cleavage mutants is identical to that seen in the wild-type virus. **C** As in B, viewed from the side, through the centre of the pentamer. Figure originally produced by me for publication in Mattei et al. (2018) and reproduced here. See Publications and Permissions section for details.

We next attempted to detect the presence of pentamers in each of the mature lattices observed in the MA-CA, CA-NC and CA-SP1 cleavage mutants, using the approach previously described by Mattei et al. (2016a) for the wild-type mature CA lattice. Only two pentamer positions were determined within the MA-CA viruses, which was insufficient to determine a structure of the pentamer in this cleavage mutant. 14 pentamer positions were detected in the CA-NC cleavage mutant viruses and 11 in the CA-SP1 mutant, at which subtomograms were extracted and averaged with C5 symmetry applied in order to obtain low-resolution structures of these pentamers. The resolution of these pentamer maps was estimated by computing the Fourier shell correlation (FSC) of each map using the 0.5 criterion, against a previously published wild-type pentamer structure obtained from authentic viruses (EMDB accession number: EMD-3466) (Mattei et al., 2016a). The

pentamer structures obtained in this way from the CA-NC and CA-SP1 cleavage mutants were identical to the wild-type pentamer, to the resolution obtained (Figure 2.8).

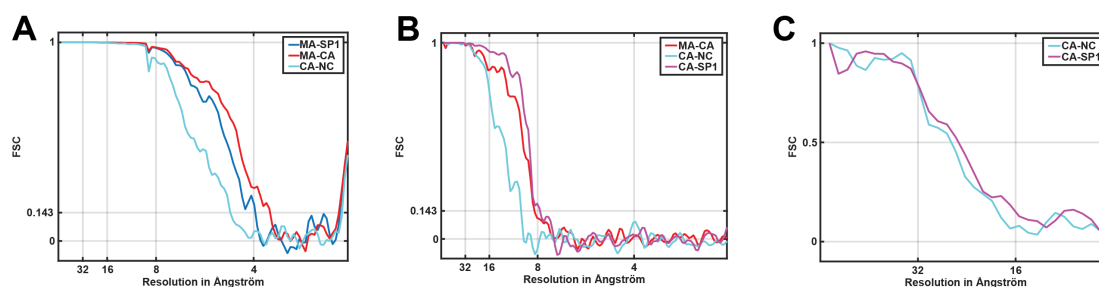


Figure 2.9: Fourier shell correlation (FSC) for immature and mature CA hexamer and pentamer structures in all cleavage mutants. **A** Gold-standard FSC for immature MA-SP1, MA-CA and CA-NC hexamer maps. **B** Gold-standard FSC for mature MA-CA, CA-NC and CA-SP1 hexamer maps. **C** Cross FSC between the wild-type pentamer structure (EMDB accession number: EMD-3466) and the pentamer structures determined from the CA-NC and CA-SP1 cleavage mutants, respectively.

2.5 Discussion

2.5.1 The complex relationship between proteolytic and structural maturation

HIV-1 maturation is a complex process which involves rearrangement of viral components at multiple levels in order to produce wild-type, infectious viruses containing a closed, mature capsid. Proteolytic maturation must occur correctly at the five cleavage sites in Gag in order to free its component domains and allow them to undergo structural maturation, in which the CA domain rearranges and forms different inter- and intra-hexamer interactions to assemble a mature lattice which has very different packing from the immature lattice. Architectural maturation is also required, in which the truncated, spherical immature lattice is remodelled into a mature capsid with fullerene cone geometry, closed by exactly 12 pentamers.

Both structural and architectural maturation are very sensitive to the degree of proteolytic processing at each cleavage site as well as the kinetics of cleavage, and numerous studies have found that blocking or changing the rate of cleavage results in the formation of defective viruses which exhibit aberrant capsids that are incomplete, are irregular in shape or do not contain the genome (Accola et al., 1998; Checkley et al., 2010; de Marco et al., 2012; de Marco et al., 2010; Göttliger et al., 1989; Kräusslich et al., 1995; Lee et al., 2009; Müller et al., 2009; Wiegers et al., 1998). However, the link between successful cleavage at each site in Gag and the structural changes that occur in the Gag lattice as a result was still unclear. In this chapter, I present direct structural evidence that addresses

this question and builds upon these previous studies in understanding the link between proteolytic, structural and architectural maturation.

A key finding here is that no single proteolytic cleavage is absolutely required for structural maturation to occur, and instead, structural maturation can be achieved using different combinations of cleavages around the CA-SP1 module. The use of large data sets has enabled the detection of minority phenotypes which show that as long as cleavage occurs on either end of CA, as is the case in the MA-CA and CA-NC mutants, structural maturation can occur albeit at a low rate. This is in contrast to the previous findings of de Marco et al. (2010) which suggested that cleavage on both ends of the CA-SP1 module was required for maturation to occur. However, the cleavage state is not the only factor in successful assembly of a wild-type capsid. The order and kinetics of cleavage also play a key role, and although structural maturation was possible in some of the cleavage mutants here, architectural maturation was always defective. This suggests that the relationship between proteolytic, structural and architectural maturation is complex and dependent upon multiple factors in order for the virus to achieve infectivity.

2.5.2 The beta-hairpin is dispensable for maturation

The high-resolution structures obtained of the immature and mature CA lattices in these cleavage mutants allow the structural changes that occur in viruses with different proteolytic cleavage states to be assessed on the level of secondary structure. One surprising result from this is that the immature Gag lattice within CA-NC contains an extra density corresponding to a beta-hairpin, typically only found in the mature CA lattice. Following the publication of the first near-atomic resolution structure of the immature lattice from within authentic HIV-1 viruses (Schur et al., 2016), there has been speculation that the immature lattice could theoretically accommodate the beta-hairpin, but evidence in support of this was lacking (Mattei et al., 2016b). The immature CA-NC Gag lattice structure here shows that cleavage between MA and CA can, and does, result in beta-hairpin formation, but that this does not destabilise the lattice sufficiently to allow efficient maturation.

In the MA-CA mutant, which is complementary to CA-NC as cleavage is blocked on the MA end of CA but allowed on the NC end, the mature CA hexamer structure lacks a beta-hairpin density which is seen in the other mature cleavage mutants as well as in the wild-type mature virus (Mattei et al., 2016a). This structure clearly shows that a mature CA lattice can form in the absence of a beta-hairpin, which cannot form in the MA-CA mutant as the linker between MA and CA is uncleaved. This is in agreement with *in vitro* experiments in which deletion of the first 12 CA residues that form the beta-hairpin still allows assembly of mature CA lattices (Gross et al., 1998), but the structure presented here

is the first evidence of this within actual viral particles. Taken together, the structures of the immature lattice in the CA-NC mutant and the mature lattice in MA-CA show that the beta-hairpin is dispensable for maturation. These results are therefore incompatible with the previously proposed role of beta-hairpin formation as a structural switch controlling maturation (von Schwedler et al., 1998).

2.5.3 CA-SP1 bundle unfolding is the structural switch for maturation

The structures of the immature Gag lattice in the MA-SP1, MA-CA and CA-NC cleavage mutants also show that the CA-SP1 six-helix bundle is always present in the immature lattice regardless of the proteolytic cleavage state of Gag. Conversely, ordered density corresponding to the CA-SP1 junction is notably absent in all of the cleavage mutant constructs in which maturation was detected, including the CA-NC mutant which is not cleaved at this site. These observations are consistent with the structures of mature CA-SP1 assemblies *in vitro*, in which the CA-SP1 junction is disordered and has been shown by NMR spectroscopy to be predominantly non-helical (Li et al., 2000; Wang et al., 2017). Destabilisation of the six-helix bundle at the CA-SP1 junction is therefore correlated with structural maturation and still occurs even when the cleavage site between CA and SP1 cannot be cleaved. This suggests that the six-helix bundle is incompatible with the mature lattice and that its dissolution is necessary for structural maturation.

2.5.4 A model for HIV-1 maturation

It is currently debated as to whether architectural maturation occurs via complete disassembly of the immature lattice followed by assembly of a mature capsid (Keller et al., 2013), or by an in-place displacive transition (Frank et al., 2015). A completely displacive model for architectural maturation (Frank et al., 2015) is, however, incompatible with formation of a closed capsid without strain defects and both experimental evidence from cryo-ET of authentic viruses and computational simulations of capsid formation favour the disassembly-reassembly model, or a mixture of disassembly-reassembly and displacive transition (Ning et al., 2016; Schur et al., 2016). It should be noted that the phenotypes seen in the mature CA-NC and MA-CA cleavage mutant viruses here, in which immature lattices are seen in close contact to mature lattices in the same virus, are not necessarily evidence in support of a displacive model of capsid formation. These phenotypes may simply be due to the remaining MA-membrane or NC-RNA tethers causing the two types of lattice to remain in place. The partially mature MA-CA and CA-NC cleav-

age mutants are reminiscent of HIV-1 particles incubated with the maturation inhibitor PF-46396 (Keller et al., 2013), and these observations suggest that HIV-1 maturation is processive, with maturation of one hexamer destabilising the local lattice and causing neighbouring hexamers to be more likely to mature. However, this is compatible with either the disassembly-reassembly model or the displacive transition model for architectural remodelling of the CA protein layer.

Mature CA lattice formation involves very different stabilising interactions from those found in immature Gag hexamers, and thus the immature lattice needs to be destabilised in order for maturation to proceed. Immature Gag lattice stability is maintained by a number of different interactions, both within and between immature CA hexamers as well as formation of a six-helix bundle at the CA-SP1 junction (Pak et al., 2017; Schur et al., 2016; Wagner et al., 2016), interactions between MA and the viral envelope (Wieggers et al., 1998) and tethering effects from RNA bound to the NC domain (Kutluay et al., 2014). Formation of an infectious, wild-type virus involves the resolution of these interactions, allowing CA to rearrange into a closed, mature capsid around the RNP genome. The results of this analysis of Gag cleavage mutants suggest that the MA-membrane and NC-RNA interactions have a cumulative effect in stabilising the immature lattice, as cleavage on either end of CA that removes either set of interactions allows maturation at low levels, and cleavage on both ends results in efficient structural maturation.

Formation of a beta-hairpin does not appear to have enough of a destabilising effect on the immature lattice to allow maturation, nor is it required for mature lattice formation, as shown by its presence in immature CA-NC lattices and absence in mature MA-CA lattices. In contrast, CA-SP1 bundle destabilisation is always correlated with structural maturation, and maturation inhibitors such as bevirimat are known to prevent maturation by stabilising the helical conformation of this region in the immature lattice (Schur et al., 2016; Wagner et al., 2016). This agrees with the fact that resistance mutations to maturation inhibitors destabilise the CA-SP1 bundle, thus overcoming the stabilising effect of these compounds on the immature lattice (Adamson et al., 2006; Schur et al., 2016; Wagner et al., 2016; Waki et al., 2012). I therefore suggest that CA-SP1 helical bundle destabilisation, rather than beta-hairpin formation, is the structural switch for HIV-1 maturation. The helical form of the CA-SP1 junction is maintained by cumulative interactions from other Gag domains, namely MA-membrane and NC-RNA tethers, and as such no single cleavage is essential for maturation. Removal of some of these interactions allows for low levels of maturation, and removal of all of them allows structural maturation to proceed efficiently, with cleavage at the CA-SP1 site then making the change permanent. Initiation of maturation at such sites then occurs in a processive manner, allowing further structural maturation of the surrounding regions of the immature lattice and architectural maturation via one of the previously proposed models discussed above.

Immature HIV-1 Gag assembly and initiation of CA maturation

Viral assembly, budding and maturation are tightly-linked processes, and disrupting any of these steps can give rise to defects in maturation and formation of non-infectious viral particles. Despite the importance of viral assembly in determining infectivity, many details of the immature Gag lattice assembly process are still unclear. For example, the identity of the basic unit of lattice assembly and the roles of different inter- and intra-domain interactions in the overall stability of the immature lattice are still unclear.

The mechanism of lattice disassembly and how this leads to initiation of maturation are also unanswered questions. As shown in the previous chapter, unfolding of the CA-SP1 helical bundle in immature Gag is required for structural maturation to occur. The rate-limiting proteolytic cleavage site in the maturation process resides within this helical bundle, which hinders proteolytic cleavage and requires bundle unfolding to proceed. The exact mechanism for protease access to this site and the spatial localisation of this process within the context of the immature Gag lattice are still not fully known.

In this chapter, I present the results of subtomogram classification experiments on a cryo-ET data set of purified, immature HIV-1 viral particles, and show that the immature Gag lattice contains partial hexamers at its edges, lacking between one and three Gag molecules. The pattern of conserved interactions in these structures strongly suggests that Gag dimers are the basic unit of lattice assembly. The CA-SP1 bundle is surprisingly ordered in these partial hexamers and appears to be more prone to destabilisation than in complete bundles, suggesting a mechanism for initiation of maturation at these sites.

3.1 Introduction

3.1.1 Immature HIV-1 Gag lattice assembly

As introduced in Chapter 1, HIV-1 assembly is driven by the Gag polyprotein which polymerises at the plasma membrane of an infected cell and recruits other viral components to the site of assembly, inducing membrane curvature and eventually recruiting the components of the host ESCRT system in order to trigger budding of an immature viral particle. Coarse-grained (CG) molecular dynamics (MD) simulations of immature Gag lattice assembly suggest that the formation of immature CA domain hexamers is mainly driven by formation of a six-helix bundle at the CA-SP1 junction (Pak et al., 2017), which I showed in Chapter 2 to be the main structural switch controlling viral maturation.

The amphipathic helix formed by the CA-SP1 junction exposes hydrophobic side chains that impose hexagonal order on the CA domains, by assembly of a helical bundle via the classical knobs-in-holes packing seen in coiled coils (Wagner et al., 2016). These relatively weak interactions give rise to a very dynamic, reversible assembly process that prevents the assembling lattice from becoming trapped in kinetically unfavourable states (Pak et al., 2017), as is the case with assembly of icosahedral viruses (Hagan et al., 2011; Rapaport, 2008). It is not surprising, therefore, that the energetics of this process are tightly controlled and highly dependent on scaffolding effects from the viral RNA as well as local curvature induced by the MA domain of Gag in order to ensure productive viral assembly (Pak et al., 2017; Yang et al., 2018).

Fluorescence microscopy experiments indicate that the HIV-1 Gag lattice assembles at the plasma membrane from low-order multimers such as dimers or trimers, rather than from larger pre-assembled Gag oligomers from the cytoplasm of the host cell (Ivanchenko et al., 2009; Kutluay & Bieniasz, 2010). This is supported by analysis of the diffusion pattern of fluorescently-labelled Gag, which shows that the viral RNA is a crucial factor in the assembly of low-order multimers of Gag into higher-order oligomers at the membrane (Yang et al., 2018). However, the identity of the low-order Gag multimer species that forms the basic unit of immature lattice assembly is still unclear, although HIV-1 Gag exists in a monomer-dimer equilibrium in solution (Rosé et al., 1992) and existing evidence in the literature suggests that the unit of assembly is likely to be either a dimer or a trimer (Accola et al., 2000; Crist et al., 2009; Yang et al., 2018).

Molecular dynamics simulations of Gag multimerisation at assembly sites suggest that it is possible for immature lattices to contain kinetically frustrated assemblies of Gag that represent incompletely-assembled hexamers lacking one or more Gag subunits (Pak et al.,

2017). However, direct experimental evidence is still required to corroborate the existence of incomplete Gag hexamers as found in these simulations. This may have important implications considering that the immature HIV-1 Gag lattice forms an incomplete spherical shell containing small, random defects and a large gap, giving rise to discontinuous edges in a lattice of hexamers stabilised by both inter- and intra-hexamer interactions (Schur et al., 2016). Although the structure of the Gag hexamer has been determined from within authentic viruses as well as virus-like particles of assembled Gag to near-atomic resolution (Schur et al., 2016), these structures obtained by subtomogram averaging represent an average hexamer with a full complement of neighbours exhibiting hexagonal packing within the lattice. The edges of the immature lattice at defects and gaps remain to be structurally characterised and are of great interest, as determining the structure of Gag species at lattice edges can help to address the question of the multimeric state of Gag that forms the basic assembly unit and also enable an assessment of the relative importance of the different interactions involved in stabilising the immature hexamer.

Here, I implemented an algorithm previously described by Heumann et al. (2011), in the Briggs group in-house subtomogram averaging pipeline, and then used this to identify and classify hexamers at the edges of Gag lattices from a previously published data set of immature HIV-1 viral particles (Schur et al., 2016). I present structures of immature hexamers from lattice edges that lack 1, 2 or 3 Gag subunits which show that Gag dimers are likely to be the basic unit of lattice assembly. I also show that the CA-SP1 helical bundle remains ordered in such incomplete hexamers, which is supported by molecular dynamics simulations performed in collaboration with Alexander Pak. Together, these results also suggest a possible role of partial hexamers at lattice edges in the initiation of Gag structural maturation.

3.2 Aims

The work in this chapter was performed in order to answer the following questions:

1. What is the structure of the Gag hexamer at the edges of the immature lattice, given that both intra- and inter-hexamer interactions are involved in stabilising the hexamer? What interactions are still present at the edges of the lattice?
2. Can subtomogram classification be used to obtain structures of immature lattice edge hexamers to a sufficient resolution to determine the interactions important in stabilising the edge of the lattice?
3. If there are incomplete hexamers of Gag at the edge of the lattice, which parts of

the CA domain remain ordered in these hexamers?

4. What can be inferred about Gag lattice assembly from the pattern of interactions in hexamers at the edges of the immature lattice? What multimeric form of Gag is the basic unit of lattice assembly?

3.3 Materials and Methods

3.3.1 Cryo-ET data set of purified immature HIV-1 viruses

The acquisition and processing of the cryo-ET data described in this section was carried out by Florian Schur and published in Schur et al. (2016). I have subsequently obtained the reconstructed tomograms and aligned subtomogram positions from Schur et al. to be used as input for my classification workflow, described in subsequent sections.

A cryo-ET data set of purified, immature HIV-1 viral particles, originally produced by Schur et al. (2016) and which yielded a 4.2 Å structure of the immature hexamer (EMDB accession number: EMD-4017), was used as a starting point for analysis of immature Gag lattice edges. This data set consists of 74 tomograms containing 484 viruses previously used for structural determination, with an unbinned pixel size of 1.35 Å/pixel. The following is a summary of the protocol used by Schur et al. (2016) to generate the tomograms in this data set:

Tilt series images were manually aligned using 10 nm gold fiducials and per-tilt defocus estimation was performed using CTFFIND4 (Rohou & Grigorieff, 2015). The contrast transfer function (CTF) was corrected for by phase-flipping using *ctfphaseflip* from the IMOD package (Xiong et al., 2009), followed by tomogram reconstruction by weighted back-projection in IMOD. Tomograms were serially binned with anti-aliasing applied in order to obtain volumes binned by factors of 2, 4 and 8. See Schur et al. (2016) for a detailed description of viral production and purification, as well cryo-ET data set acquisition parameters and the subtomogram alignment workflow used to generate the published in-virus high-resolution structure of immature Gag.

3.3.2 Removal of duplicate and misaligned subtomogram positions

Roughly-aligned subtomogram positions from Schur et al. were obtained from an intermediate step in the processing of the data set above, and were generated by three succes-

sive iterations of alignment and averaging of $8 \times$ binned subtomograms against an initial 6-fold symmetric reference, as described in Schur et al. (2016). The points were chosen from this iteration in order to ensure that the subtomogram alignment had converged on a hexagonal Gag lattice, but that no subtomograms had yet been removed from the data set based on low cross-correlation coefficient values.

As subtomogram extraction positions were oversampled for the initial angular search, duplicate subtomograms that had aligned onto the same positions were removed from the data set by applying a pairwise distance criterion of 4 binned pixels (4.32 nm). The remaining positions were then visualised in UCSF Chimera using a custom plugin as described in Qu et al. (2018), and misaligned subtomogram positions were removed by manual inspection. Misaligned positions were defined as those positions not conforming to the geometry of the hexagonal lattice, for example those that were rotated more than 90° out of plane relative to the lattice. This cleaning step was performed manually in order to avoid point removal by cross-correlation coefficient (CCC), which could result in potential lattice edges being removed from the data set if, for example, if an edge subtomogram did not correlate as well to the 6-fold hexameric reference as an average non-edge hexamer would.

3.3.3 Initial geometric selection of lattice edges

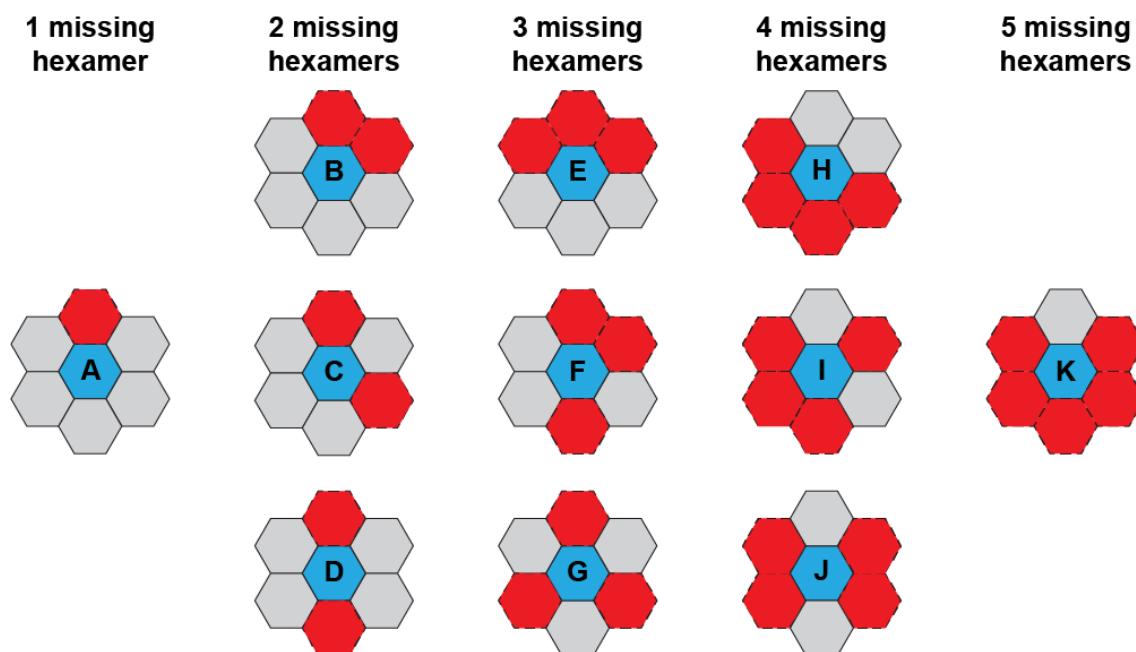


Figure 3.1: Possible combinations of missing neighbours in a hexagonal lattice. The hexamer being classified based on its neighbours is depicted in light blue and is labelled A-F depending on which class it belongs to. Neighbouring hexamers that are present are shown in grey and missing hexameric neighbours are represented by red, dashed hexagons.

The remaining 178 750 aligned subtomogram positions were then used as a starting point to identify the edges of the immature Gag lattice. In order to identify candidate subtomograms for further image-based classification, I wrote a MATLAB script to identify every possible pattern of missing neighbours around each hexamer position in the lattice map of each virus in the data set (Figure 3.1). Using this script, I identified 62 815 potential edge positions and oriented all of them so that the predicted gap in the lattice would always point in a single direction. This was done to enable averaging of the points to obtain an initial structure.

As the number of subtomograms was very low in the non-contiguous gap classes (Figure 3.1 C-D, F-G, I-J) as well as all classes missing 4 or 5 neighbours, (Figure 3.1 H, K) I only retained the subtomograms classified as having a pattern of 1, 2 or 3 contiguous neighbouring hexamers missing (Figure 3.1 A-B, E). This resulted in a data set of subtomogram positions containing 57 134 points, which I subsequently pooled for use with image-based classification, described in the following sections.

3.3.4 Further alignment prior to classification

The coordinates of the oriented edge hexamer subtomogram positions identified by geometric analysis of lattice maps were then scaled for use with $4\times$ binned data. Subtomograms were extracted from $4\times$ binned tomograms with a box edge size of 72 binned pixels, corresponding to 388.8 \AA in each dimension. One iteration of fine alignment was then performed against the final $4\times$ binned average previously generated by Schur et al. (2016) as part of their original structural determination of the immature hexamer from this data set. The alignment was performed using an $8 \times 2^\circ$ angular search range for all Euler angles, a 32.4 \AA low pass filter, C6 symmetry and a mask around the central hexamer and all six neighbouring positions. This was done in order to improve the quality of the aligned positions before subsequent image-based classification of the subtomograms, and the resulting subtomogram positions were later used as the starting point for both image-based classification approaches described in the following section. The average of the aligned subtomograms was also generated for subsequent use in wedge-masked difference map and multi-reference alignment based classification.

3.3.5 Image-based subtomogram classification approaches

Principal component analysis (PCA) of wedge-masked difference maps (WMD)

In order to approach the problem of classifying lattice edge hexamers, I implemented wedge-masked difference map-based subtomogram classification (hereafter referred to as WMD PCA), originally described by Heumann et al. (2011), in the in-house subtomogram averaging software package developed in the Briggs group based on the TOM (Nickell et al., 2005), AV3 (Förster et al., 2005) and Dynamo (Castaño-Díez et al., 2012) packages. In this approach, the rotations and shifts previously found by aligning the subtomograms to a 6-fold symmetric hexamer reference were then applied to each subtomogram to produce rotated subtomogram volumes in register with the reference volume. The average of all the correctly rotated and shifted subtomograms was also generated as explained in the previous section.

The approach used to generate the Fourier weights in the following text was originally devised by Dustin Morado at the MRC Laboratory of Molecular Biology, and was used by me as part of my implementation of wedge-masked difference based classification.

Fourier weights of the same box size (388.8 \AA^3) were then generated in order to compensate for the missing wedge and CTF of the tomogram from which each subtomogram was extracted. This was done by using the subtomogram positions to define empty regions in the corresponding tomograms for extraction of volumes containing noise but no Gag protein. Subtomograms containing noise were extracted with the above box size at 100 positions per tomogram and normalised to a mean grey value of 0 with a variance of 1. For each tomogram, the amplitude spectra of the corresponding noise subtomograms were calculated by computing the square root of the power spectrum of each subtomogram, and these amplitude spectra were averaged to produce a Fourier weight for that tomogram that describes missing information in Fourier space due to the missing wedge and the CTF.

End of text describing the use of an approach for generating Fourier weights devised by Dustin Morado.

Wedge-masked difference maps were then calculated for each of the rotated subtomograms described above, as follows. For each rotated subtomogram, the same set of rotations was applied to the Fourier weight for the corresponding tomogram to generate a correctly rotated wedge mask. The rotated subtomogram as well as the average of all rotated subtomograms were normalised to a mean grey value of 0 and variance of 1 and low pass filtered to a resolution of 29.9 \AA . The computed wedge mask was then applied to both of these volumes by calculating the Fourier transform of the subtomogram and the aver-

age, element-wise division by the wedge mask followed by an inverse Fourier transform to generate the wedge-weighted volumes. A mask was then applied to the weighted volumes so that only the central hexamer and its six immediate neighbours were considered for difference map calculation. The grey values under the mask were again normalised to zero mean and unit variance, and the weighted and masked subtomogram volume was subtracted from the weighted and masked average of all rotated subtomograms in order to generate a wedge-masked difference map.

The difference map voxels under the masked region of interest were then stored as an $m \times n$ matrix, where m is the number of voxels under the mask and n is the number of subtomograms in the data set. Singular value decomposition (SVD) was then performed on the matrix of difference map voxels to decompose the voxel matrix D according to the relationship $D = USV^T$. The first 30 left singular vectors of the matrix were obtained from the matrix U , reshaped to match the size of the region of interest mask applied and stored as the first 30 eigenvolumes of the data set. SV^T was stored as this provides the corresponding eigencoefficients for use in clustering the data.

The eigenvolumes were inspected manually in order to determine which ones corresponded to structural differences between the average structure and the subtomograms, and which ones mainly described differences in orientation of the subtomograms relative to the missing wedge, and a subset of good eigenvolumes was selected. The eigencoefficients corresponding to these selected eigenvolumes were used as input for k -means clustering in MATLAB with 10 replicates and $k = 30$. The subtomograms in the data set were then grouped into classes based on these clusters, and the average of each class was generated from the same $4 \times$ binned subtomograms used before. Class averages were inspected visually, and classes containing 1, 2, 3 as well as no missing hexameric neighbours around the central hexamer were identified, with some classes rotated in-plane by 1 hexamer position (i.e. 60°). The in-plane rotation angle of the subtomogram positions in these rotated classes was adjusted in order to match the configurations seen in the other classes. Multiple classes were identified as missing 1, 2 and 3 neighbouring hexamers, and these classes were pooled into larger, single classes with 1, 2 and 3 missing neighbouring hexamers before generation of the class averages.

Multi-reference alignment and classification

The aligned, $4 \times$ binned lattice edge subtomograms were used as a starting point for the construction of artificial references to be used in multi-reference alignment and classification of the subtomograms. The subtomogram positions were divided into equal-sized subsets according to odd and even particle number and averaged in order to produce two starting

averages, each with half of the data. This was done in order to allow for oversampling of each class during multi-reference alignment and classification by making use of the random variation between the two half-references. The odd and even half-references were then multiplied in MATLAB by masks constructed in order to down-weight the positions in each reference corresponding to 1, 2 or 3 missing hexamers around the central hexamer (Figure 3.2). The original, non-multiplied references were also included to allow for classification of subtomograms that may actually contain complete hexamers with a full complement of 6 neighbours.

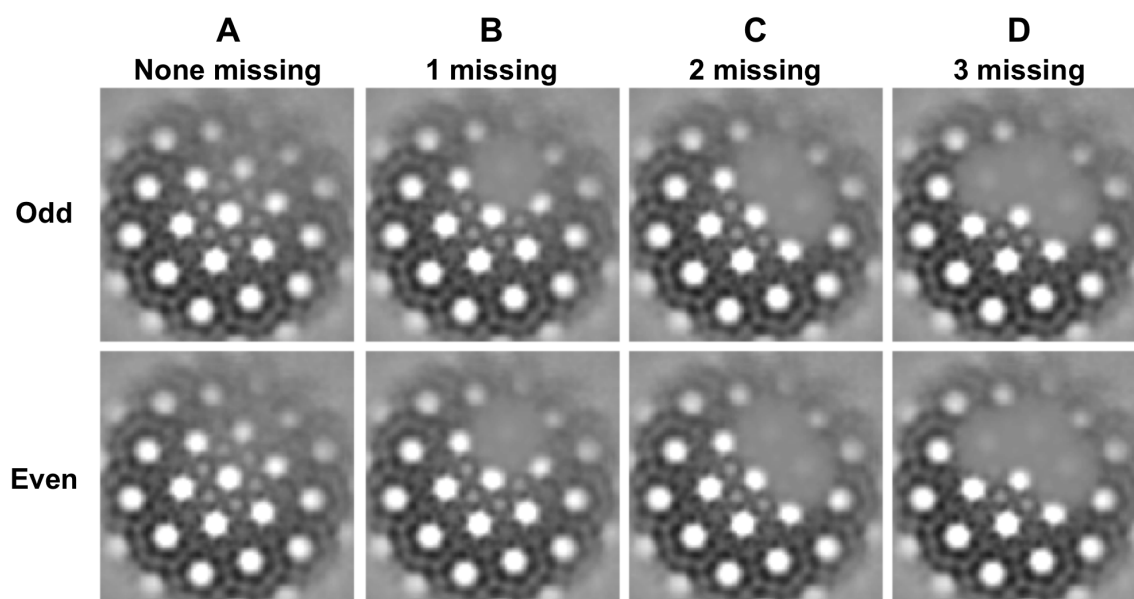


Figure 3.2: Central XY slice through the references constructed for multi-reference subtomogram classification. Two references were constructed for each case, using either the odd-numbered or even-numbered half of the data set, respectively. **A** Averages of the odd and even halves of the subtomogram data set to be classified. **B** Artificial references for the case with 1 missing neighbour, made by down-weighting part of the density in **A**. **C** Artificial references for the case with 2 missing neighbours, made by down-weighting part of the density in **A**. **D** Artificial references for the case with 3 missing neighbours, made by down-weighting part of the density in **A**.

The subtomograms were then aligned against each of these artificial references for 6 iterations, with a low pass filter of 29.9 Å, a mask around the central hexamer as well as three of its contiguous neighbours in the gap direction used for geometric classification and a restricted in-plane angular search allowing only rotation of 60° in each direction in order to correctly classify subtomograms that had been misclassified by up to one neighbour during geometric classification earlier. The reference to which each subtomogram aligned with the highest cross-correlation coefficient CCC was used to assign the class of that subtomogram. Simulated annealing was used for stochastic sampling in order to allow subtomograms to escape from local minima between alignment iterations, using a scaling factor of 0.3 with the approach described by Hrabe et al. (2012). Class membership had converged onto stable classes by the sixth iteration, and the results of this alignment iteration were used for subsequent structure generation and analysis.

Unbinning subtomogram classes to generate final structures

The pooled classes from the WMD PCA classification approach described above, as well as the control classes from multi-reference classification, were then unbinned to regenerate the class averages from $2\times$ binned subtomograms by scaling the coordinates by a factor of 2. Subtomograms were extracted from $2\times$ binned tomograms with a box edge size of 128 binned pixels, corresponding to 345.6 Å in each dimension, and averaged to generate the new references. Further unbinning to non-binned data was also attempted in order to improve the resolution of the classes, but the resolution was not measurably better as assessed by gold-standard FSC using a criterion of 0.143. The references generated using $2\times$ binned data were therefore used as the final structures.

3.3.6 Molecular dynamics simulations

This section describes molecular dynamics simulations performed by Alexander Pak (University of Chicago, Illinois, United States of America) in collaboration with me, based on my structures obtained by subtomogram averaging.

In order to investigate the properties of partial helical bundles as seen in the incomplete hexamer structures lacking between 1 and 3 Gag subunits, molecular dynamics simulations of the CA-SP1 junction in these bundles were carried out with restraints applied to helix 9 of the CA-CTD in order to emulate a lattice. All-atom molecular dynamics simulations of the solvated CA-SP1 junction of immature HIV-1 Gag protein were performed multiple times with different parameters to simulate different components in the buffer, including Tris, ethylenediaminetetraacetic acid (EDTA) and inositol hexakisphosphate (IP₆), in order to account for possible differences due to the rings of lysine residues shown by Dick et al. (2018) to coordinate IP₆. The free-energy landscape of each incomplete and complete helical bundle was sampled using metadynamics and time-structure based independent component analysis (tICA), in order to simulate the paths for folding and refolding of the CA-SP1 junction in each case.

End of section describing molecular dynamics simulations carried out by Alexander Pak.

3.4 Results

3.4.1 Classification of Gag hexamer species at the lattice edge

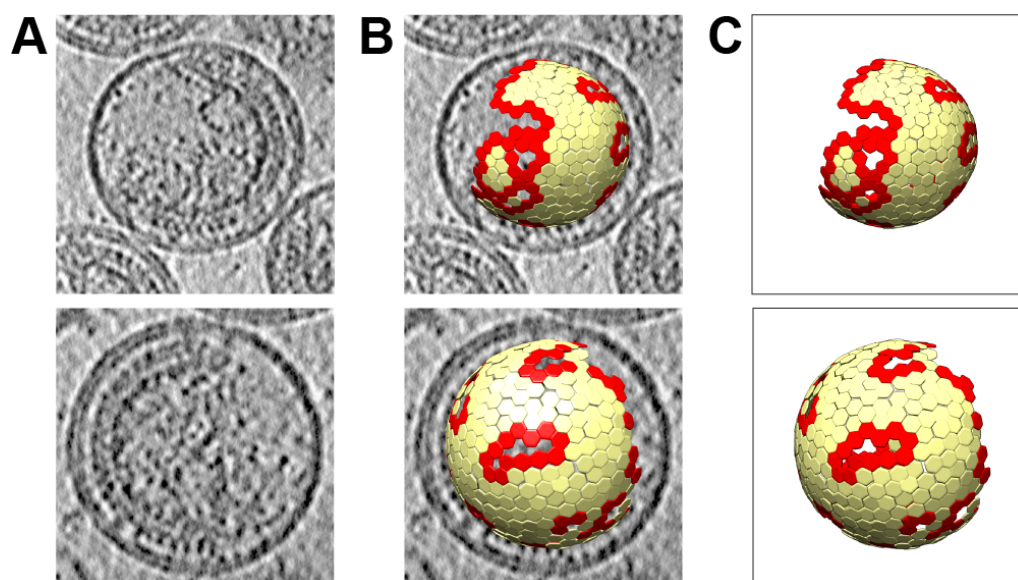


Figure 3.3: Computational slices through an example tomogram with overlaid subtomogram positions. **A** Computational slices of 5.4 Å thickness through a 4× binned, CTF-corrected and Gaussian-filtered (5 pixel kernel) tomogram showing examples of the immature viruses used for Gag lattice edge classification. The thick density under the viral membrane corresponds to the immature Gag lattice and a large gap is visible in which there is no assembled Gag polypeptide. **B** As in A, with overlaid lattice maps to show the aligned subtomogram positions and orientations obtained from the initial reference-free alignment on 8× binned data performed by Schur et al. (2016). Misaligned subtomogram positions were manually removed from the data as described in Materials and Methods. Gag hexamers at the edges of the lattice (red) were identified geometrically using a MATLAB script that I wrote to classify and orient edge hexamers based on the local configuration of neighbours around each hexamer. Non-edge hexamers are shown in yellow. The hexamers shown in red form a pattern of small defects and a single, large defect. **C** As in B, but without the corresponding slice from the tomogram.

The tomogram shown in this figure was generated by Florian Schur at the European Molecular Biology Laboratory (EMBL), Meyerhofstraße 1, 69117 Heidelberg, Germany.

I used a published cryo-ET dataset, from which Schur et al. (2016) determined a 4.2 Å map of immature HIV-1 Gag directly from purified viruses (EMDB accession number: EMD-4017), as a starting point for analysis of hexamers at the edges of the immature lattice (Figure 3.3). Roughly-aligned coordinates of hexamer subtomograms were visually inspected and obviously misaligned points, defined as those which did not exhibit hexagonal packing, were manually removed from the data set. Fine angular alignments were then performed using 8× binned and then 4× binned data. The refined coordinates were then geometrically analysed using a custom MATLAB script in order to identify discontinuous lattice edges. These points were used as input for further image classification of the hexamer subtomograms.

Image classification of subtomogram data to reveal differences in macromolecular struc-

ture is complicated by the missing wedge problem (Bartesaghi et al., 2008; Förster et al., 2008; Heumann et al., 2011), which describes missing information in Fourier space due to physical limitations on the angular range across which a sample can be tilted in the electron microscope. The impact of this missing information on classification results tends to be much larger than real structural differences in the complex of interest, which can result in clustering of subtomograms based on orientation relative to the optical axis unless computational methods are used to compensate for this missing information (Förster et al., 2008).

In order to overcome this, I implemented wedge-masked difference map based classification (WMD PCA) (Heumann et al., 2011) in the Briggs group in-house subtomogram averaging pipeline. This, in addition to better estimation of the missing wedge and CTF by using the average amplitude spectrum of noise in each tomogram (Dustin Morado, unpublished - See Materials and Methods), allowed for good separation of structural classes while ignoring missing wedge orientation differences. I then used this approach on the data set of subtomograms that had previously been identified and oriented using geometric classification (see Materials and Methods for details). Singular value decomposition (SVD) of the matrix of difference map voxels allowed me to determine eigenvolumes that describe the variation in the difference maps computed under the mask (Figure 3.4). I then visually selected eigenvolumes that appeared to contain significant density in neighbouring hexamer regions, corresponding to structural variability in those regions, and classified the subtomograms based on the corresponding eigencoefficients by *k*-means clustering in order to generate structural classes (Figure 3.5). The results of WMD PCA classification of the data set were also validated by performing an independent multi-reference alignment and classification using synthetic references. The details of both of these classification approaches are described in more detail in Materials and Methods.

Although class membership was not identical between the two classification approaches when applied to the same input data, this is not unusual for classification of noisy, missing wedge-affected data, and both approaches still converged on exactly the same structural classes with only minor differences in how clean the separation was between classes. Both of these classification approaches were able to identify separate populations of immature hexamers at the edge of the lattice, with 1, 2 or 3 neighbouring hexamers missing. These structures provide views of the structure of the lattice edge for the first time (Figure 3.5).

From the structural classes obtained, it can be clearly seen that for each missing neighbouring hexamer, there is a corresponding loss of one Gag subunit density in the central hexamer at that position (Figure 3.5), with structural classes identified which lack 1, 2 or 3 Gag subunits from the central hexamer. These structures are the first direct evidence of incomplete Gag hexamers at the edges of the immature lattice. Hexamers lacking 4 or 5

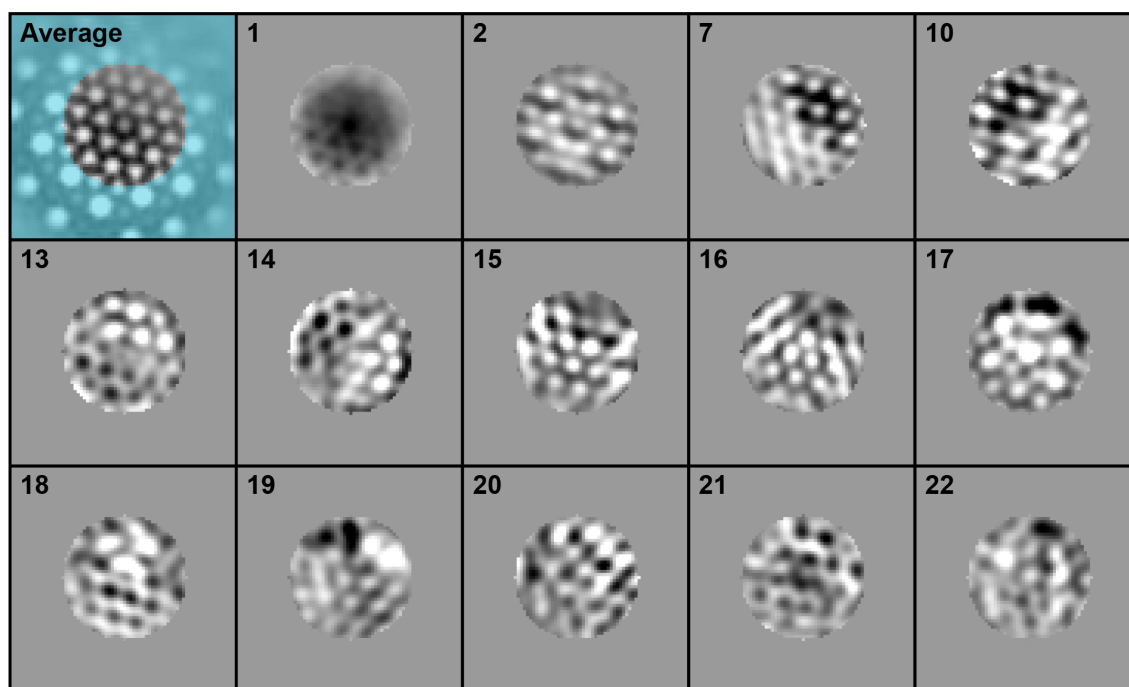


Figure 3.4: Eigenvolumes selected for clustering of the subtomogram data set based on wedge-masked difference maps. XY slices at the CA-CTD level through the average of all subtomograms, and each of the eigenvolumes chosen for k -means clustering of the data set, are shown. The average volume used for difference map calculation is overlaid with the mask marking the region of interest, with regions in cyan multiplied by zero and not considered.

Gag subunits were not identified by classification, which could imply one of the following possibilities: 1) hexamer species lacking 4 or 5 subunits may not exist at the edge of the lattice, and it is difficult to imagine where they would fit on the edges seen in this data set, or that 2) these species may exist but either do not align to a hexameric reference or classify out because the lack of so many subunits makes them very dissimilar from a complete hexamer. It was also difficult to map out the spatial arrangement of defects along the edge by classification due to differences in class membership.

3.4.2 Dimer and trimer interfaces in incomplete Gag hexamers

The partial hexamer structures obtained, which are missing 1, 2 or 3 Gag molecules, also shed light on the relative importance of the different CA-CA interactions in maintaining lattice stability. In these structures, when an edge hexamer lacks one or more CA-CTD, electron density corresponding to the dimer partner of that domain is also missing. This shows that dimers are either present or absent as a single unit, and isolated monomers were not observed in the partial hexamer structures obtained. I also observed that when one CA-CTD dimer was missing from a hexamer, the density for two CA-NTD trimers was also missing. These missing CA-NTD densities correspond to the Gag position for which dimer density was lost as well as the CA-NTD of the Gag molecule one position

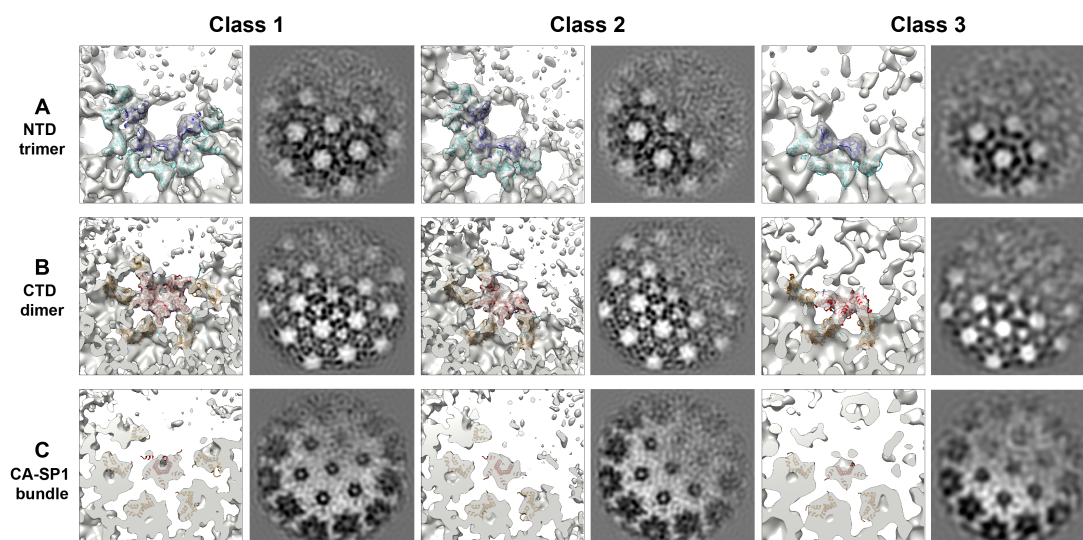


Figure 3.5: Structures of incomplete hexamers at immature Gag lattice edges obtained by WMD PCA classification. **A** Isosurface (left panel) and orthoslice (right panel) representations of the density corresponding to the CA-NTD layer in each of the structural classes obtained by WMD PCA subtomogram classification. Classes 1, 2 and 3 represent Gag hexamers that are missing density corresponding to 1, 2 and 3 contiguous neighbouring hexamers, respectively, and it can be seen that these classes correspond to the same number of missing Gag molecules from the central hexamer. The NTD from an atomic model of the immature hexamer (PDB accession code: 5L93) is shown fit into each individual Gag position in the central hexamer (dark blue) and the corresponding CA-NTDs that form NTD trimers with the central hexamer (cyan). For each missing neighbouring hexamer, the corresponding NTD trimer density is absent as well as the neighbouring trimer immediately counter-clockwise from it. **B** As in A, but viewed at the level of the CA-CTD dimeric interface. The CA-CTD from the same model as in A is shown fit into the dimeric interfaces formed by the protomers from the central hexamer (red) with those of the neighbouring hexamers in the lattice (orange). There is clear loss of one CTD dimer density for each missing hexameric neighbour. **C** As in B, but viewed at the level of the CA-SP1 six-helix bundle. The six helix bundle density is surprisingly still ordered to some degree even upon removal of 1, 2 or 3 Gag molecules from the central hexamer.

anti-clockwise from the missing CA-CTD, viewed from the direction of the CA-NTD towards the CA-CTD. In the assembled lattice, the CA-NTD forms a trimeric interface, and these data suggest that when the trimer is lacking one member, the remaining NTDs are no longer stabilised in their positions in the lattice (Figure 3.5).

The pattern of interfaces seen in the structures of the immature hexamer shows that the dimer interface is always maintained and that there are no detectable instances in which the Gag subunit in central hexamer remains ordered without this dimeric interface. The pattern of dimer and trimer interfaces seen therefore suggests the importance of the CTD dimer interface in stabilising the NTD of immature CA, providing evidence for the idea that the basic assembly unit of the immature lattice is the Gag dimer as opposed to the Gag trimer. This is consistent with reports in the literature showing that minimal Gag constructs containing only the myristoylation site on MA and the CA-CTD, but not the CA-NTD, are assembly competent as long as inositol hexakisphosphate IP₆, an abundant cellular polyanion selectively packaged into immature HIV-1 particles, is also present

(Accola et al., 2000). Molecular dynamics simulations of the assembly properties of the CA-CTD with the CA-SP1 junction included also lend support to this idea, suggesting that given scaffolding effects from membrane and RNA attachment, the CA-CTD and SP1 helical bundle interfaces are sufficient to allow assembly in the absence of CA-NTD trimeric interfaces (Pak et al., 2017).

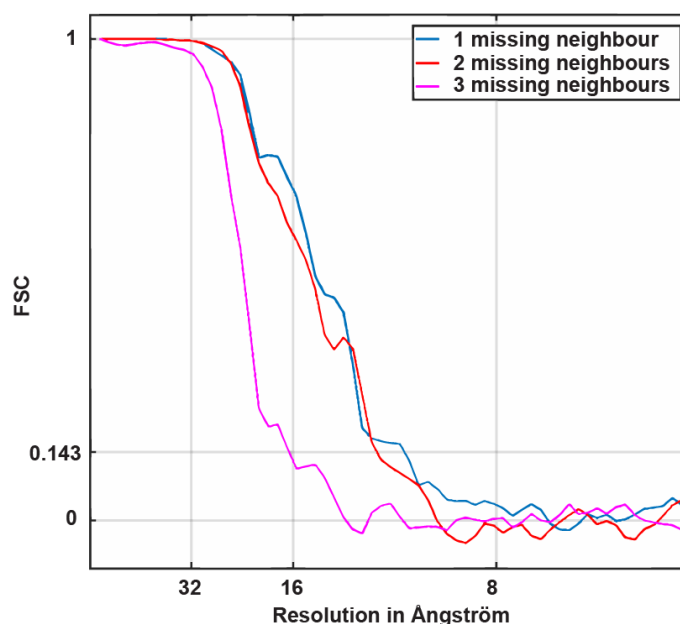


Figure 3.6: Fourier shell correlation (FSC) of final WMD PCA incomplete hexamer classes, with 1, 2 or 3 missing Gag hexamer neighbours. The structures were generated by averaging $2\times$ binned subtomograms as the use of unbinned data did not improve the resolution obtained. The resolution obtained for each of the structures was as follows: 1 missing neighbour, 11.2 Å; 2 missing neighbours, 11.1 Å; 3 missing neighbours, 16.2 Å.

3.4.3 CA-SP1 junction integrity in incomplete hexamers

The density maps of incomplete lattice edge hexamers also show that when one or more Gag subunits are missing from a hexamer, the CA-SP1 six-helix bundle density does not completely disappear from the maps (Figures 3.5 and 3.7). This suggests that the CA-SP1 helical bundle can still exist in a stable form even when fewer than 6 helices are present, and that the bundle structure can adapt to loss of a helix without becoming completely disordered.

This raises the question of how the bundle accommodates loss of up to half of its constituent helices while retaining ordered packing, given that a crystal structure of this region in a full CA-SP1 six-helix bundle exhibits classical knobs-in-holes packing of the hydrophobic residues exposed along the amphipathic CA-SP1 helix (Wagner et al., 2016). In these structures, the helical bundle density becomes slightly weaker with increasing numbers of missing Gag subunits from the central hexamer, but the resolution

of the maps is insufficient to unambiguously place individual helices within the bundle structure in hexamers missing 1, 2 and 3 Gag subunits (Figure 3.7). Additionally, the density corresponding to the loop between helices 10 and 11 in the CTD was observed to be weaker when the neighbouring CTD in the hexamer was absent. This suggests that local structural changes may occur next to defects in incomplete hexamers, such as increased flexibility in the part of the CTD directly upstream of a partial helical bundle (Figure 3.7).

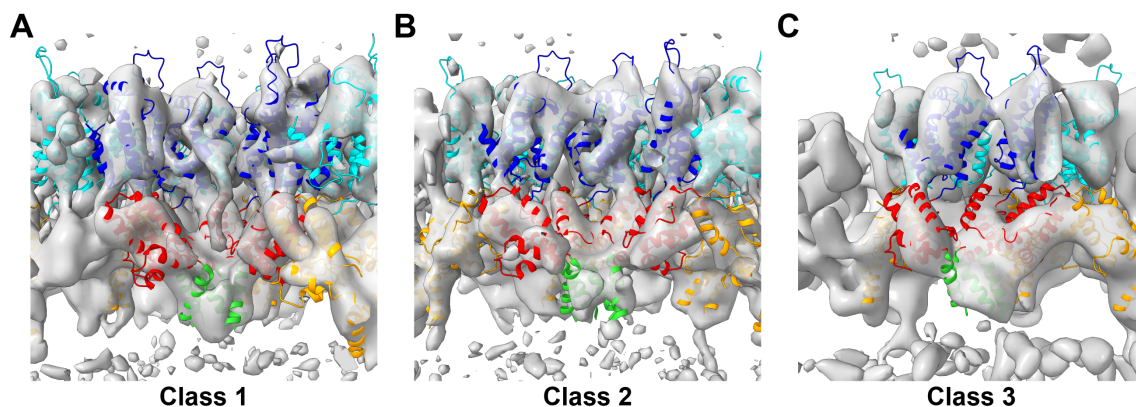


Figure 3.7: Side view of incomplete hexamer maps from the direction of the missing neighbouring hexamers, showing an isosurface render of the maps with **A** one, **B** two and **C** three missing Gag subunits from the central hexamer. Individual Gag subunits from an atomic model of the immature hexamer (PDB accession code: 5L93) were fit as rigid bodies into each position in the central, incomplete hexamer (blue: NTD, red: CTD) and neighbouring hexamers (cyan: NTD, red: CTD). The amino acids corresponding to the CA-SP1 junction in the central hexamer are highlighted in green. In each case, ordered bundle density is visible in this region, even though up to 3 Gag subunits are missing from the central hexamer. The density in the CA-CTD that corresponds to helices 10 and 11 on one side of the defect in the incomplete hexamer is also visibly weaker than the rest of the density, as shown by the poor fit in the region.

The following text describes the results of molecular dynamics simulations carried out in collaboration with Alexander Pak, based on the cryo-EM density maps that I obtained by subtomogram classification.

In order to further investigate the dynamics of the CA-SP1 helical bundle region in both complete and incomplete hexamers, we performed molecular dynamics simulations on this region. In all cases where the hexamer was missing 1 to 3 subunits, the position of the CA-SP1 junction (helix 12) was observed to shift on average, although we conclude that the overall conformation of this region remained helical, within the sampling limitations of our simulations. A previous study using nuclear magnetic resonance (NMR) spectroscopy to characterise the dynamics of the CA-SP1 junction in hexameric assemblies of immature Gag have shown that this region in complete hexamers exists in a state of helix-coil equilibrium (Wang et al., 2017). Extending this finding to incomplete hexamers, we conclude from our simulations that the helix-coil transition in an incomplete CA-SP1 helical bundle has many intermediate states, that is to say multiple free-energy barriers, rather than a simple A→B transition between the two states. These simulations

showed that the CA-SP1 helix adjacent to the gap in the hexamer, in the counter-clockwise direction as viewed from the NTD towards the CTD, was more likely to become disordered than the clockwise gap-adjacent helix which had a transition barrier similar to that calculated for helices in a complete six-helix bundle.

End of text describing molecular dynamics simulations by Alexander Pak.

Interpreting my incomplete hexamer structures in the context of the molecular dynamics simulations described above, I therefore conclude the partial helical bundles formed by the CA-SP1 junction are ordered on average, although they may be more prone to destabilisation than helices in a complete bundle. This has potentially important implications for the initiation of maturation in the context of the results presented in Chapter 2. This is discussed in more detail in the following sections.

3.5 Discussion

3.5.1 Gag dimers are likely to be the basic immature lattice assembly unit

The structures I obtained here by subtomogram classification directly show the existence of partial Gag hexamer species at discontinuous edges of the immature lattice, revealing the arrangement of Gag at the edges formed when lattice assembly has stopped and budding has occurred as well as the interfaces that are preserved when inter-hexamer interactions cannot form at such edges. It is important to note that these structures most likely do not represent transient intermediates, but stable end states, as they have been obtained from viruses that have budded and have subsequently undergone an extensive purification process. This is supported by published evidence from molecular dynamics simulations of immature Gag lattice assembly, in which kinetically frustrated assemblies of Gag at lattice defects lead to incorporation of incomplete hexamers at edges of the lattice (Pak et al., 2017).

The pattern of interfaces within and between the remaining hexamers at a lattice defect allows the relative importance of these interfaces to be inferred based on structures of incomplete hexamers obtained by classification of hexamers at these edges (Figure 3.8). These structures show that the absence of one neighbouring hexamer, which removes the CA-CTD dimer partner for one of the Gag molecules in the central hexamer, also results in loss of density for that dimer. In the same case, the density corresponding to two NTD trimers adjacent to the missing hexameric neighbour was also missing from the map. In

the NTD, the removal of one neighbouring hexamer makes one protomer from each of the two adjacent NTD trimers unavailable to form trimeric interactions. These results show that partial trimers of NTD, lacking one subunit but with the other two subunits in the same position, are not ordered when interactions with the neighbouring hexamer are abolished. Presence or absence of ordered CTD density at the dimeric interface in the central hexamer was also seen to depend on the presence of the dimer partner from the corresponding adjacent hexamer.

It should be noted, however, that this does not require the neighbouring hexamer position which contributes a dimer or trimer partner to the central hexamer to itself be complete. Indeed, this is unlikely to be the case as the pattern of defects observed in immature HIV-1 Gag lattices (Figure 3.3) would necessitate tiling of incomplete hexamers lacking varying numbers of Gag molecules in order to satisfy the pattern of interfaces seen. Considering the pattern of dimeric and trimeric interactions seen in incomplete hexamers, this suggests that whole hexamers of Gag are unlikely to be the multimeric form in which Gag is added to a growing immature lattice. Similarly, a trimer is also unlikely to be the basic unit of lattice assembly, as in this case the pattern of missing trimers from an incomplete hexamer would be expected to correlate with the number of missing neighbours while the pattern of CTD dimer interactions would be considerably more variable. When both the NTD trimer and CTD dimer patterns observed in the maps are considered, the multimeric species of Gag most likely to fulfill the role of basic assembly unit is the Gag dimer.

These observations are consistent with the fact that Gag naturally exists in equilibrium between the monomeric and dimeric state in solution (Rosé et al., 1992), and that leucine-zipper constructs that induce Gag dimerisation can assemble correctly if inositol hexakisphosphate (IP₆) is also present (Crist et al., 2009), which is the case in wild-type viruses given the importance of IP₆ in both the immature (Dick et al., 2018) and mature (Mallery et al., 2018) stages of the viral life cycle. This is also supported by the observation that a minimal Gag construct containing only the myristoylation site on MA (matrix), the CA-CTD and SP1 is assembly competent even though the NTD is not present to form trimers (Accola et al., 2000). The requirement for IP₆ in dimer-forming leucine zipper constructs and the requirement for SP1 in constructs lacking the CA-NTD is consistent with molecular dynamics simulations suggesting that multimerisation of Gag within an assembling lattice is driven by CA-SP1 helical bundle formation (Pak et al., 2017), the importance of which in maturation has been shown in Chapter 2 of this thesis.

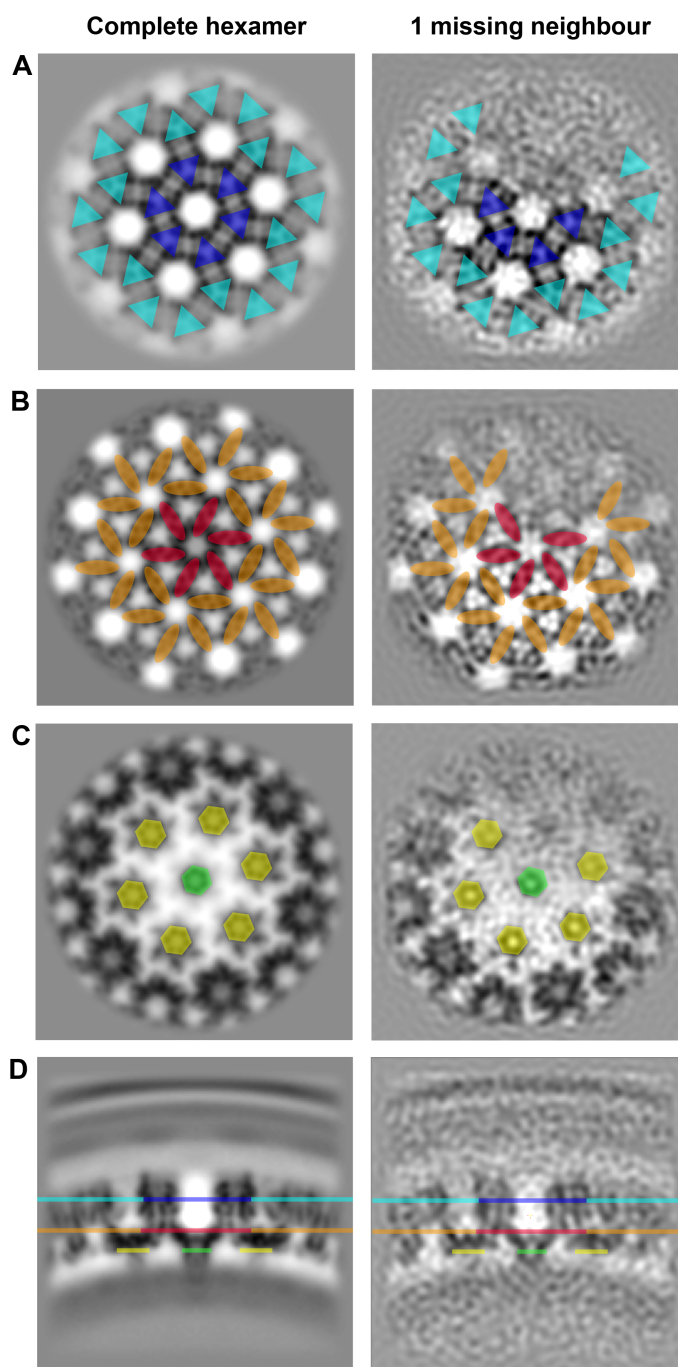


Figure 3.8: Schematics of the interactions stabilising the immature HIV-1 Gag lattice. Each panel shows a computational slice through a structure of a complete, 6-fold symmetric immature hexamer generated from $2\times$ binned data by Schur et al. (2016) during the subtomogram averaging workflow that they used to generate their published 4.2 \AA structure of the immature hexamer from within authentic virions (EMDB accession number: EMD-4017), as well as the corresponding slice in the hexamer structure lacking one neighbour that I determined. The slices are overlaid with the following geometric objects to highlight the interfaces involved in stabilising the immature hexameric lattice, and those that are lost when one neighbouring hexamer is absent: **A** triangles representing CA-NTD trimers (blue: central hexamer, cyan: neighbouring hexamers), **B** ellipses representing CA-CTD dimers (red: central hexamer, orange: neighbouring hexamers) and **C** hexagons representing the CA-SP1 helical bundle (green: central hexamer, yellow: neighbouring hexamers). Panel **D** shows a side view through the Gag lattice with overlaid rectangles to show the approximate positions of the highlighted layers coloured as in panels A-C. **The complete hexamer density map shown in this figure was generated by Florian Schur at the European Molecular Biology Laboratory (EMBL), Meyerhofstraße 1, 69117 Heidelberg, Germany.**

3.5.2 The CA-SP1 region still forms partial helical bundles in incomplete hexamers

A surprising observation from these incomplete hexamer structures was that the CA-SP1 helical bundle was still ordered even when up to 3 Gag subunits were missing from a hexamer. This shows that the CA-SP1 six-helix bundle can accommodate loss of one or more Gag molecules, but the resolution of the maps is insufficient to unambiguously place the remaining helices within the density seen. This raises questions about the integrity and dynamics of partial helical bundles, and in order to address these, I collaborated with Alexander Pak who performed molecular dynamics simulations of the CA-SP1 region in incomplete hexamers. We found that partial helical bundles can remain ordered on average, but there is an increased probability of uncoiling exhibited by the CA-SP1 region of the gap-adjacent Gag molecules. This effect appears to be directional, with the Gag position directly anti-clockwise from the missing Gag molecule position as viewed from the NTD towards the CTD, being the most prone to CA-SP1 destabilisation.

The observation that the helical bundle is still ordered in incomplete hexamers and is more likely to be destabilised than a full six-helix bundle has potentially important implications for the initiation of maturation. As I showed in Chapter 2, unfolding of the CA-SP1 helical bundle appears to be the main structural determinant of the transition from the immature Gag lattice to a mature CA lattice. However, where along the immature lattice maturation is first initiated and how maturation events continue across the rest of the immature lattice are still open questions. The six-helix bundle formed by the CA-SP1 junction has been shown to exist in a helix-coil equilibrium, and it is this equilibrium that is thought to allow maturation events and subsequent protease access to the rate-limiting cleavage site within the bundle (Wang et al., 2017) (also see Chapter 2). Increased propensity of partial helical bundles towards the uncoiled state would mean an increased likelihood of maturation at incomplete hexamers relative to complete ones as well as an increased probability of undergoing proteolytic cleavage between CA and SP1, due to the higher probability of the bundle becoming disordered.

As discussed in Chapter 2, there are two competing models for how the immature lattice is remodelled into a mature lattice with completely different stabilising interactions and different hexamer spacing: 1) disassembly of the immature lattice into small oligomers or multimers followed by construction of a mature lattice (Keller et al., 2013), and 2) an in-place phase transition that does not involve disassembly (Frank et al., 2015). If incomplete hexamers are indeed preferential sites of maturation initiation, it is unclear how a hexamer lacking up to 3 Gag subunits could be remodelled by an in-place phase transition into a mature CA hexamer or pentamer, which would have to be complete in

order to form a closed capsid. This may add support to the hypothesis that immature lattice disassembly is required, at least to some degree, in order to form a properly closed capsid during maturation. A revised model incorporating the results presented here would involve stochastic initiation of maturation at the edges of the lattice due to the higher probability of CA-SP1 unfolding in incomplete hexamers, which would then cause local disassembly of the immature lattice as it undergoes structural maturation. This, in turn, would destabilise hexamers immediately adjacent to the maturation event by removing the inter-hexamer interactions that are involved in both dimer and trimer formation in the immature lattice, and these hexamers would then be more likely to undergo maturation compared to those in the middle of the lattice. This would proceed towards the middle of the lattice as a ‘wave’ of maturation from one or more initiation sites.

3.5.3 Stability of the immature lattice in the context of retroviral evolution

It is also informative to consider the arrangement of interactions seen in incomplete hexamers of immature HIV-1 Gag in the wider context of retroviral evolution. Identification of the CTD dimer as the basic unit of immature lattice assembly in HIV-1 is consistent with the fact that there is a high level of tertiary and quaternary structural conservation of the CA-CTD, but not the CA-NTD, across different retroviral families, as seen from structures of Gag in immature Rous sarcoma virus (Bharat et al., 2012), murine leukemia virus (Qu et al., 2018), Mason-Pfizer monkey virus and HIV-1 (Schur et al., 2015). This is further evidence of importance of the CTD dimer interface in maintaining the stability of the immature lattice, which is consistent with the observation that mutations such as WM184,185AA in the CA-CTD dimer interface of HIV-1 Gag result in defective viral assembly (Joshi et al., 2006; Ono et al., 2005). This is seen in the incomplete hexamer structures presented here, in which loss of a CTD dimer interface always results in the loss of the corresponding trimer interactions, but loss of the NTD trimer interface in a given position in the hexamer is still compatible with an ordered, immature lattice provided the corresponding CTD dimer interface is intact.

The tertiary and quaternary structure of the CA-CTD in the mature hexamer is even more conserved, with extraordinarily high similarity not only between retroviral families (Mattei et al., 2016a; Qu et al., 2018), but also between HIV-1 and much more distantly related retrotransposons such as Ty3/Gypsy (Dodonova et al., 2019) and the neuronal retrotransposon Arc (Erlendsson et al., 2019) despite very low sequence conservation between these different lineages. This is in contrast to the immature forms of various retroviruses, which have more divergent quaternary structure in the CA-NTD but relatively similar CA-CTD

quaternary arrangement (Qu et al., 2018). This divergence between immature retroviral Gag structures has led to suggestions that the immature form of Gag is a relatively recent evolutionary adaptation to an extracellular stage in the viral life cycle, enabled by acquisition of domains such as an envelope protein and a matrix domain (Dodonova et al., 2019). The immature form of the virus would therefore be the assembly-competent form of the virus, whereas the mature virus is the infectious form with a capsid that protects the viral genome and mediates nuclear entry, in the case of HIV-1. It is tempting to speculate that incomplete hexamers of immature HIV-1 Gag are a way for the transition between these two forms to be effected, by facilitating maturation via destabilisation of partial CA-SP1 helical bundles.

Modulation of HIV-1 capsid stability and nuclear import by host cofactors

The mature HIV-1 capsid serves many different functions during the different stages of infection. It must be stable enough to maintain a closed structure and prevent degradation of the viral RNA genome by host nucleases, but at the nucleus, uncoating must be triggered in order to facilitate nuclear import of the viral pre-integration complex (PIC).

The conical HIV-1 capsid presents a variety of potential binding sites for small molecule and protein cofactors, and there is extensive evidence that host cofactors are important in regulating HIV-1 nuclear import and capsid stability. There has been particular interest in the interaction of the capsid with inositol hexakisphosphate (IP₆), which both promotes immature Gag assembly and greatly increases the stability of mature HIV-1 capsids.

In this chapter, I present near-atomic resolution structures I determined from conical *in vitro* assemblies of HIV-1 CA in the presence of IP₆ by subtomogram averaging. I also present the highest resolution structure of the viral pentamer to date, which has long eluded crystallisation attempts, opening up the possibility of using this system for routine structural characterisation of potential pentamer-binding cofactors.

I then used these CA assemblies as a platform to determine high-resolution structures of CA hexamers and pentamers upon addition of peptides containing the known CA-binding sequences of nucleoporin 153 (Nup153) and cleavage and polyadenylation specific factor 6 (CPSF6), factors crucial for nuclear import which share the same binding pocket but have a different mode of binding. This analysis shows that both of these cofactors bind to CA hexamers, but show no detectable binding to pentamers at the same concentration.

4.1 Introduction

The HIV-1 capsid performs a multitude of different functions upon entry into the cell. The capsid is not only a physical barrier against immune recognition and degradation of the viral genome (Lahaye et al., 2013; Rasaiyaah et al., 2013), but also serves to import nucleotides for reverse transcription of the RNA genome into proviral DNA (Jacques et al., 2016) and then mediate import of its contents across the nuclear membrane and target it to integration sites in transcriptionally-active chromatin of the host genome (Burdick et al., 2017; Sowd et al., 2016). The mechanical stability of the capsid is critical to these roles, as the capsid must first be trafficked intact to the nuclear membrane in order to prevent degradation of the viral genome, but must uncoat readily enough to then perform its function in importing the pre-integration complex (PIC) into the nucleus.

Many questions about the assembly, stability and uncoating of the capsid remain unanswered and various approaches have been taken to dissect these processes, from coarse-grained molecular dynamics simulations (Grime et al., 2016) to fluorescent assays that allow capsid uncoating to be monitored in real-time within permeabilised viruses, avoiding the need for extensive capsid purification protocols (Márquez et al., 2018). Capsid uncoating assays by monitoring loss of fluorescence (Márquez et al., 2018) have shown that approximately 60–80 % of capsids are incomplete and contain defects, whereas this number was reported to be even higher by cryo-electron tomography of capsids within purified, authentic viruses (Mattei et al., 2016a). This observation is seemingly at odds with the protective functions of the capsid, especially as capsid disassembly seems to occur rapidly as soon as a defect is present in the CA lattice, and these defective capsids appear to only be stable within the virus due to the pool of free CA protein at high concentration within the viral envelope (Márquez et al., 2018).

The surface of the HIV-1 capsid, which consists of a variable number of CA hexamers that tile to form a fullerene cone, provides many different potential binding interfaces for small molecule and host cofactors, as the lattice has to adopt significantly different curvatures along the capsid and also incorporate CA pentamers, which are now known to have appreciably different CA-CA interfaces instead of being strictly quasi-equivalent to the hexamer (Mattei et al., 2016a). Interaction of the capsid with various cofactors has been found to significantly change its properties, from stability within the cytosol (Mallery et al., 2018) and mediation of trafficking towards the nucleus (Huang et al., 2019), to recognition by the nuclear pore and subsequent import of its contents (Di Nunzio et al., 2012).

4.1.1 Inositol hexakisphosphate (IP₆)

One cofactor that has been the subject of particular attention in recent years is inositol hexakisphosphate (IP₆), a small molecule that is abundant in mammalian cells at concentrations of approximately 10–20 μ M (Letcher et al., 2008). IP₆ is also important in assembly of the immature virus, where it binds to a six-helix bundle between the C-terminal end of CA and the first part of SP1, and is an assembly cofactor and determinant of correct viral assembly (Campbell et al., 2001; Dick et al., 2018). In mature viruses, IP₆ also plays an important role as a pocket factor that greatly increases the stability of mature capsids against spontaneous disassembly, and also increases DNA synthesis within the capsid by a factor of over 100 compared to when it is absent (Mallery et al., 2018). IP₆ binds to a different site in the mature CA hexamer, where it interacts with a conserved ring of six arginine residues (R18) below the beta-hairpin domain (Mallery et al., 2018), which is the same electrostatic pore proposed by Jacques et al. (2016) to mediate import of nucleotides into the capsid for reverse transcription.

It is currently unknown whether IP₆ interacts with the CA pentamer as structural characterisation of IP₆ bound to CA has so far only been carried out by X-ray crystallography of cross-linked hexameric CA. Attempts to use similar disulfide cross-linking methods to stabilise a pentamer have given rise to a CA pentamer that is quasi-equivalent to the hexamer and formed using the same interactions (Pornillos et al., 2011), but it is now known that the CA pentamer formed in authentic viruses is significantly different in structure as it forms using different inter-protomer interfaces (Mattei et al., 2016a). It has therefore not been possible to use crystallography to investigate whether IP₆ also binds to the pentamer and whether this has a physiological function, but it is possible that this does occur as the structure of the pentamer from authentic viruses has a strong density in the centre of the equivalent R18 pore (Mattei et al., 2016a), which in the hexamer corresponds to IP₆ (Mallery et al., 2018). Viral pentamers are selectively incorporated into the capsid lattice at the broad and narrow ends which exhibit higher curvature than the rest of the capsid (Mattei et al., 2016a), raising the question of whether they may play a role in capsid stabilisation and destabilisation by IP₆ and other cofactors.

4.1.2 Nuclear import cofactors

Interaction of host cofactors with the HIV-1 capsid plays a crucial role in the mediation of nuclear import. These cofactors include nuclear-localised components such as cleavage and polyadenylation specific factor 6 (CPSF6) (Lee et al., 2010; Rasaiyaah et al., 2013) and cyclophilin A (CypA) (Franke et al., 1994) as well as components of the nu-

clear pore, namely nucleoporin 153 (Nup153) and nucleoporin 358 (Nup358, otherwise known as RANBP2) (Di Nunzio et al., 2012). CPSF6 and Nup153, in particular, are interesting as they have been suggested to regulate capsid uncoating as part of their nuclear import function (Márquez et al., 2018). These two cofactors perform different functions during infection, with capsid mutation experiments showing that Nup153 is required for nuclear translocation in non-dividing cells whereas CPSF6 is important in viral genome integration (Buffone et al., 2018).

Nup153 and CPSF6 share a common binding site formed between the N-terminal domain of one CA protomer and both the N-terminal and C-terminal domains of the adjacent CA molecule within the same hexamer, although there are some differences in the specific residues that bind to either of the two cofactors (Price et al., 2014). At the core of this binding site is a hydrophobic pocket formed by side chains from helices 3, 4 and 5 as well as the linker between helices 5 and 6 of the CA N-terminal domain, which is bound to by a conserved phenylalanine residue in both Nup153 and CPSF6 (Price et al., 2014). Nup153 and CPSF6 also form other interactions with the neighbouring CA molecule in the hexamer, with Nup153 mainly interacting with the N-terminal domain and CPSF6 mainly interacting with C-terminal domain residues instead (Price et al., 2014). This common binding interface is also shared with small molecule antiretroviral drugs that include PF-3540074 (PF74) (Blair et al., 2010) and BI-2 (Lamorte et al., 2013), with PF74 being of particular interest due to its apparent ability to both stabilise the capsid lattice and induce defect formation at concentrations above 10 μ M (Márquez et al., 2018; Shi et al., 2011). Small molecule compounds from the GS-CA family including GS-CA1 and GS-6207 have also been reported by Gilead Sciences, Inc. to bind to the same pocket as PF74, CPSF6 and Nup153, although the crystal structures supporting this are currently not publicly available (Tse et al., 2017). These compounds have generated much interest as antiretroviral drug candidates due to their potency (Singh et al., 2019; Zheng et al., 2018).

As with IP₆, structural characterisation of Nup153 and CPSF6 binding to HIV-1 CA has been limited to crystallographic studies using cross-linked CA hexamers, which do not provide information on the localisation of particular modes of binding along the curved capsid surface. Subtomogram averaging of the CA hexamer from authentic viral cores has been successful in preserving this spatial information via classification of lattice curvature (Mattei et al., 2016a), but the limited resolution of mature CA hexamers resolved in this way has precluded analysis of a possible dependence of cofactor binding on lattice curvature. Due to the elusiveness of the viral CA pentamer as a target for crystallography, it is still unknown whether viral pentamers have a specific binding preference to particular viral cofactors, if they are involved in recognition of the capsid by the nuclear pore and whether Nup153 and CPSF6 bind to the viral pentamer at all. These questions are

important as the exact mechanism of how capsid uncoating is initiated or prevented, and whether specific parts of the capsid are recognised by the nuclear pore based on curvature are unknown.

4.1.3 *In vitro* conical assemblies of HIV-1 CA with IP₆

Here, I present structural work that I have performed on *in vitro* virus-like particles (VLPs) consisting of recombinant HIV-1 CA assembled with IP₆, prepared by my collaborator Robert Dick, which I show contain both hexamers and pentamers that are identical to those in the native viral capsid. Using this system, I have obtained near-atomic resolution structures of the mature CA hexamer in various curvature states as well as a significantly higher resolution structure of the viral pentamer than previously possible. I then used these *in vitro* assemblies of CA to investigate the occupancy of IP₆ in different parts of the capsid, including pentamers. I also added peptides containing the known sequences of Nup153 and CPSF6 that interact with their common binding interface on the capsid, and determined high-resolution structures of CA hexamers and pentamers from VLPs incubated with these peptides in order to investigate whether they bind to the pentamer and whether they exhibit binding specificity for particular parts of the capsid.

4.2 Aims

I performed the work in this chapter in order to address the following questions:

1. Is the structure of the mature CA hexamer, and more importantly the pentamer, the same in these *in vitro* conical assemblies as in the wild-type virus, and if so, can these VLPs be used to improve the resolution achievable by subtomogram averaging of the viral form of the pentamer?
2. Does IP₆ bind to the viral form of the CA pentamer as it does to the hexamer?
3. Does the more open binding site in the viral form of the pentamer still support binding of Nup153 and CPSF6, and if so, is there a difference between binding to the hexamer and pentamer?
4. Does lattice curvature in the capsid alter binding of Nup153 and CPSF6 to their common binding pocket, and is there preferential binding of these cofactors to particular regions of the capsid?

4.3 Materials and Methods

4.3.1 Expression, purification and in vitro assembly of conical CA cores with IP₆

This section was originally written by my collaborator, Robert Dick, for a manuscript in preparation, and has been reproduced and modified by me. Robert Dick performed the following work to express, purify and assemble HIV-1 CA into conical IP₆-stabilised virus-like particles (VLPs), at Cornell University, Ithaca, New York, United States of America.

“HIV CA was expressed and purified from *E. coli*. Cells harvested by centrifugation were resuspended in 20 mM Tris-HCl [pH 8], 2 mM TCEP (tris(2-carboxyethyl)phosphine), 2 mM PMSF (phenylmethylsulfonyl fluoride), and lysed by sonication. Cellular debris was removed by ultracentrifugation in a Beckman TLA-110 rotor at a relative centrifugal force of 417k for 45 min. Nucleic acid was removed from the supernatant by precipitation via the addition of polyethylenimine (PEI) to a final concentration of 0.03 %. Ammonium sulfate to 25 % saturation was added to the resulting supernatant, and the precipitate was resuspended in 20 mM Tris-HCL [pH 8], 2 mM TCEP, filtered through a 0.2 µm filter, and the residual ammonium sulfate removed via HiPrep Desalting (GE Healthcare). The resulting sample was concentrated to ~5 ml, and flowed over tandem cation and anion exchange columns (GE HiTrap SP and HiTrap Q). The flow-through fraction contained the CA protein at ~90 % purity. To further purify, the protein was subjected to size exclusion chromatography (SEC) on a Superdex 75 Increase 10/300 column (GE). The peak corresponding to CA was collected, concentrated to ~20–35 mg ml⁻¹ (260/280 = 0.58) in 20 mM Tris-HCl [pH 8], 2 mM TCEP, flash frozen, and stored at –80 °C. Assembly reactions were performed by warming assembly buffer (50 mM MES [pH 6.0], 30 mM NaCl, 2 mM TCEP, 2.5 mM IP₆) to 37 °C followed by the addition of CA protein to a final concentration of 500 µM. Assembly reactions were incubated at 37 °C for 10 minutes and stored at 4 °C. Virus-like particle (VLP) assemblies were diluted 1:20 in buffer (50 mM MES pH 6, 2 mM TCEP, 100 µM IP₆) prior to spotting onto EM grids for negative staining with uranyl acetate.”

Written by Robert Dick and modified by me.

4.3.2 Negative staining of VLPs for morphology experiments

5 μ l of the CA VLPs assembled in the presence of IP₆ were incubated for 7 min on the carbon side of a non-glow-discharged C-Flat CF200 Cu EM grid. The excess liquid was blotted off from the edge of the grid using a Kimcare wipe, and a 5 μ l of 2 % uranyl acetate solution was placed on the same side of the grid. After incubation for a further 7 min, the excess stain was blotted off in the same way and the grid was allowed to dry before imaging by electron microscopy. Grids were imaged in an FEI Tecnai G2 Spirit 120 keV transmission electron microscope, equipped with a LaB₆ electron source and Gatan Ultrascan 1000 XP charge-coupled device (CCD) camera.

4.3.3 Cryo-EM grid preparation

C-Flat CF-2/2-3Cu-50 grids were glow-discharged for 45 seconds with a current of 25 mA in a PELCO easiGlow glow discharger immediately before use. All samples were vitrified using a Thermo Fisher Scientific Vitrobot Mark IV, operated at 100% humidity and 18 °C.

In the case where no peptide was added to the conical CA VLPs with IP₆, the sample mixture to be plunge-frozen was made by mixing VLPs in assembly buffer with BSA-conjugated 10 nm gold fiducials in 1 \times PBS at a ratio of 8:1. For the CPSF6 and Nup153 peptide binding experiments, an extra peptide-binding step was carried out in order to prepare the sample for plunge-freezing. Conical CA VLPs with IP₆ were mixed with a volume of BSA-conjugated 10 nm gold fiducials in 1 \times PBS, calculated to give a final core:gold ratio of 8:1 after peptide addition. The peptides used were CPSF6 residues 313-327 and Nup153 residues 1407-1423, obtained from Donna Mallery (MRC Laboratory of Molecular Biology, Cambridge, United Kingdom) as a powder synthesised by Designer Biosciences. A solution of the respective peptide at 10 times the required concentration, in core assembly buffer containing 1 % DMSO, was then diluted 1:10 in this mixture of cores and gold fiducials, and incubated on ice for 15 minutes prior to use.

4 μ l of the sample mixture for plunging was applied to the carbon side of a glow-discharged grid in the humidity chamber of the Vitrobot. The sample was then manually blotted from the opposite side of the grid for 3 seconds using Whatman No. 1 filter paper, and then plunge-frozen in liquid ethane.

4.3.4 Cryo-electron tomography of VLPs

| Sample | CA VLPs + IP ₆ | CA VLPs + IP ₆ + 50 μ M Nup153 ₁₄₀₇₋₁₄₂₃ | CA VLPs + IP ₆ + 50 μ M CPSF6 ₃₁₃₋₃₂₇ |
|-----------------------------|--|---|--|
| Microscope | FEI Titan Krios | FEI Titan Krios | FEI Titan Krios |
| Voltage (keV) | 300 | 300 | 300 |
| Detector | Gatan K2-XP | Gatan K2-XP | Gatan K2-XP |
| Energy-filter | Yes | Yes | Yes |
| Slit width (eV) | 20 | 20 | 20 |
| Acquisition settings | Super-resolution mode | No | No |
| | Å/pixel | 1.379 | 1.191 |
| | Defocus range (microns) | -1.5 to -4.5 | -1.5 to -4.5 |
| | Defocus step (microns) | 0.25 | 0.50 |
| | Acquisition scheme | -60/60°, 3°, Dose-symmetric | -60/60°, 3°, Dose-symmetric |
| | Total Dose (electrons/Å²) | ~123 | ~141 |
| | Dose rate (electrons/Å²/sec) | ~3.25 | ~4.3 |
| | Frame number | 10 | 10 |
| | Tomogram number | 66 | 71 |

Table 4.1: Summary of data acquisition parameters for HIV-1 CA cores assembled with IP₆, with and without added 50 μ M Nup153 peptide (1407-1423) or CPSF6 peptide (313-327).

All tilt series were acquired on a Thermo Fisher Scientific Titan Krios 300 keV cryo-transmission electron microscope, equipped with a Gatan GIF BioQuantum post-column energy filter with a slit width of 20 eV. The tilt series for the conical CA VLPs assembled with IP₆ were recorded on a Gatan K2-XP direct electron detector in operated in counting mode, with dose-fractionation of each tilt image into 10 frames. All tilt series were acquired using a dose-symmetric tilt scheme (Hagen et al., 2017), with a tilt angle range between -60° and 60° and a 3° tilt angle step. Detailed data acquisition parameters are listed in Table 4.1.

4.3.5 Image processing

| Sample | | CA VLPs + IP ₆ | CA VLPs + IP ₆ + 50 μ M Nup153 ₁₄₀₇₋₁₄₂₃ | CA VLPs + IP ₆ + 50 μ M CPSF6 ₃₁₃₋₃₂₇ |
|--|-----------------------------------|---------------------------|---|--|
| CA hexamer | VLPs | 684 | 591 | 254 |
| | Asymmetric units Set A | 269 850 | 226 740 | 94 872 |
| | Asymmetric units Set B | 269 856 | 226 740 | 94 878 |
| | Final resolution (0.143 FSC) in Å | 3.9 | 4.8 | 6.2 |
| CA pentamer | Asymmetric units Set A | 13 110 | 10 320 | 4 560 |
| | Asymmetric units Set B | 13 110 | 10 325 | 4 560 |
| | Final resolution (0.143 FSC) in Å | 6.2 | 7.0 | 8.0 |
| Asymmetric CA hexamer next to pentamer | Asymmetric units Set A | 9 791 | 7 466 | 3 229 |
| | Asymmetric units Set B | 9 658 | 7 643 | 3 148 |
| | Final resolution (0.143 FSC) in Å | 6.5 | 7.2 | 8.1 |

Table 4.2: Summary of image processing parameters for HIV-1 CA cores assembled with IP₆, with and without added 50 μ M Nup153 peptide (1407-1423) or CPSF6 peptide (313-327).

Table 4.1 summarises the image processing parameters used for structural determination of the CA VLP data sets, with and without added CPSF6 or Nup153 peptide.

Pre-processing and tomogram reconstruction

Gain correction, motion correction and dose-reweighting, as described by Grant and Grigorieff (2015), were performed on the raw tilt series movies using alignframes from the IMOD package. The defocus and angle of astigmatism of each tilt image was estimated using ctffplotter from IMOD. Tilt series were manually aligned using gold fiducials in IMOD, and tomograms were reconstructed with 3D CTF correction by CTF multiplication and astigmatism correction using novaCTF (Turoňová et al., 2017). The recon-

structed tomograms were then serially binned by factors of 2, 4 and 8 in IMOD to produce binned tomograms for use in the initial stages of subtomogram alignment.

Segmentation of HIV-1 CA VLPs from tomograms

8× binned, non-CTF corrected tomograms were converted from MRC format to a series of TIFF images and loaded into the Ilastik software package (version 1.3.2rc1) (Sommer et al., 2011) for segmentation of the VLP cores. The tomograms were loaded into the Carving workflow of Ilastik, and filtered using the ‘dark lines (valley filter)’ option with a filter scale of 5 pixels, in order to divide each tomogram into a boundary map of supervoxels.

Each CA VLP was then segmented individually by using the Background and Object labels in Ilastik to define its enclosed volume according to the computed boundary map. The segmentations were exported from Ilastik in the Wavefront OBJ format, and the vertex coordinates of the segmented volumes were extracted for import into MATLAB. I developed a MATLAB script which was then used to convert these lists of vertices into MRC volume files that could be overlaid on the original tomograms. These volume files were used to define the subtomogram extraction points.

Hexamer subtomogram alignment and averaging

The converted segmentations were used to define subtomogram extraction points by applying a low pass filter and defining an isosurface for each volume in MATLAB, followed by selection of an appropriate isosurface threshold to define the surface of the VLP. An oversampled grid of points with 3.3 nm spacing was defined for subtomogram extraction along this surface using the TOM toolbox (Nickell et al., 2005), with an initial estimate of the appropriate Euler angles based on the normal vector to the surface at each point.

Initial reference construction: An initial reference was constructed using a 4-tomogram subset of the CA VLP dataset without any peptide cofactors added, which contained 35 VLP cores. All subtomogram alignment steps were performed using subTOM, the in-house subtomogram averaging software package developed in our group, based on the TOM (Nickell et al., 2005), AV3 (Förster et al., 2005) and Dynamo (Castaño-Díez et al., 2012) packages. This subset was chosen to evenly sample the range of defocus values across the data set. Subtomograms were extracted from 8× binned, 3D CTF corrected tomograms at the sampled positions, with a box size of 706 Å³. The extracted subtomograms were averaged to obtain an initial map for reference-free subtomogram averaging.

The subtomograms were then iteratively aligned against this initial reference map for 6 iterations, with a cone search range of $8 \times 5^\circ$, in-plane rotational search of $6 \times 5^\circ$, low pass filter of 35.3 Å and no symmetry imposed.

The average converged on a lattice of CA hexamers by the sixth iteration, after which a shift was applied to the average to centre it on the 6-fold symmetry axis of a CA hexamer. One further iteration of alignment was then run with the same parameters, to ensure that the alignment converged on the shifted box position. Subtomograms that had aligned onto the same hexamer, defined using a pairwise distance criterion of 4 pixels (44.1 Å), were removed from the data set. Misaligned subtomograms were removed by imposing a per-tomogram minimum cross-correlation coefficient CCC threshold, determined by plotting the aligned subtomogram positions and orientations in a custom plugin for UCSF Chimera (Qu et al., 2018). Two further iterations of alignment were then run with a search range of $6 \times 2^\circ$ for all Euler angles, the same low pass filter as before and C6 symmetry imposed. The average from the final iteration of alignment was used as a reference for all subsequent alignments.

Alignment of full hexamer data set: The initial hexamer reference constructed was used as a starting reference for alignment of the data sets with no peptide, 50 µM CPSF6 peptide (residues 313-327) and 50 µM Nup153 peptide (residues 1407-1423). Due to the pixel size difference between the data set without peptide and the other two data sets, the reference was scaled to the correct pixel size in MATLAB before being used to align the CPSF6 and Nup153 data sets.

Subtomograms were extracted from $8 \times$ binned, 3D CTF corrected tomograms with the centre coordinates and orientations determined from segmentation of the entire data set. A cubic box of 64 pixels in each dimension was used for this step with each data set, corresponding to an actual size of 706.0 Å for the data set with no peptide bound, and 609.8 Å for the CPSF6 and Nup153 data sets. These subtomograms were aligned against the reference for 1 iteration with a $8 \times 5^\circ$ cone search, $6 \times 5^\circ$ in-plane search, 35.3 Å low pass filter (35.3 Å for CPSF6 and Nup153) and C6 symmetry imposed. Duplicate and misaligned points were removed as described in the initial reference construction protocol, with a threshold of 55.2 Å (57.2 Å for CPSF6 and Nup153). Another iteration of alignment was then performed with the same parameters, but a finer search of $6 \times 2^\circ$ for all angles.

The aligned coordinates were then scaled from a bin factor of 8 to a bin factor of 4. The subtomograms were re-extracted from $4 \times$ binned tomograms with box dimensions of 353 Å^3 (343 Å^3 for CPSF6 and Nup153). The extracted subtomograms were averaged to produce a new reference for further alignment. Two iterations of alignment were then

performed with a search range of $5 \times 2^\circ$ for all Euler angles, C6 symmetry and a 32.0 Å low pass filter (31.2 Å for CPSF6 and Nup153). Points that had aligned onto the same positions during this iteration were removed as described above, using a pairwise distance threshold of 55.2 Å (57.2 Å for CPSF6 and Nup153).

The aligned, $4 \times$ binned subtomogram positions were scaled to the pixel size of $2 \times$ binned data and subtomograms were extracted at these positions from $2 \times$ binned tomograms, with box dimensions of 353 Å³ (343 Å³ for CPSF6 and Nup153). At this point, the data sets were split into odd- and even-numbered half-sets of equal size. The averages of each half-set were generated and all further alignments were done in parallel on these completely independent halves of each data set. One iteration of $3 \times 2^\circ$ angular search for all Euler angles was carried out, with C6 symmetry and a low pass filter of 15.3 Å for the data set with no added peptide and 15.6 Å for the CPSF6 and Nup153 data sets. One more iteration was then run with the search range for all Euler angles reduced to $3 \times 1^\circ$, and low pass filters of 11.0 Å for the data set with no peptide and 9.8 Å for the CPSF6 and Nup153 data sets.

Finally, fine alignment was performed by scaling the aligned positions to the unbinned pixel size, and subtomograms were extracted with box dimensions of 264.8 Å for the data set without added peptide, and 266.8 Å for the CPSF6 and Nup153 data sets. The first iteration of fine alignment used a $3 \times 1^\circ$ angular search for all Euler angles, C6 symmetry and a low pass filter of 8.3 Å for all data sets. The final iteration of alignment was then carried out with a $2 \times 1^\circ$ angular search, C6 symmetry and low pass filter settings of 6.6 Å, 7.6 Å and 6.7 Å for the data sets with no added peptide, CPSF6 peptide and Nup153 peptide respectively. The final resolution of each map was measured by gold-standard Fourier shell correlation (FSC) between the two half data sets.

Pentamer subtomogram alignment and averaging

Pentamer subtomogram positions were obtained by identifying patterns of 5 hexamers with pairwise distance constraints that satisfy pentamer geometry, as described in Mattei et al. (2016a). The hexamer subtomogram positions after the final iteration of alignment on $4 \times$ binned data were used for pentamer identification in this manner. The pentamer subtomograms were extracted at the identified positions from $4 \times$ binned tomograms and a box size of 353 Å for the data set without added peptide and 355 Å for the CPSF6 and Nup153 data sets. The extracted subtomograms were averaged to generate an initial reference, which was then used for an alignment iteration with $6 \times 3^\circ$ angular search for all Euler angles, C5 symmetry and a low pass filter setting of 32.1 Å for the data set with no peptide and 31.2 Å for the CPSF6 and Nup153 data sets. The in-plane rotation

angle of each subtomogram position was then incremented by 0° , 72° or 144° at random in order to eliminate the possibility of orientation bias in the structure from the geometric identification step. A new average was generated and used as the reference for a further iteration of alignment with the same parameters as before, except for a finer angular search range of $6 \times 2^\circ$. Misaligned pentamer subtomograms were removed after this step by manual inspection of the lattice maps in UCSF Chimera.

The pentamer subtomogram positions were then scaled for use with $2\times$ binned data, and divided into independent odd and even half-sets as with the hexamer subtomograms. Subtomograms were extracted from $2\times$ binned tomograms and aligned with C5 symmetry, $3 \times 2^\circ$ angular search and a low pass filter of 23.5 \AA for the peptide-free data set and 21.4 \AA for the CPSF6 and Nup153 data sets. A second iteration of alignment was then performed with C5 symmetry, $3 \times 1^\circ$ angular search and a low pass filter set to 17.7 \AA for the peptide-free data set and 17.2 \AA for the CPSF6 and Nup153 data sets.

The aligned pentamer positions were then scaled to the pixel size of the unbinned data and subtomograms were extracted from unbinned tomograms with a box size of 264.8 \AA for the peptide-free data set and 266.8 \AA for the CPSF6 and Nup153 data sets. One iteration of fine alignment was run with $3 \times 1^\circ$ angular search, C5 symmetry and a low pass filter of 13.2 \AA , 10.3 \AA and 9.5 \AA for the peptide-free, CPSF6 and Nup153 data sets respectively. A second alignment iteration was then run with a $2 \times 1^\circ$ angular search range, C5 symmetry and a low pass filter of 10.6 \AA , 10.3 \AA and 8.6 \AA for the peptide-free, CPSF6 and Nup153 data sets respectively. The final resolution of each map was measured by gold-standard FSC between the two half data sets.

Identification and alignment of hexamers next to pentamers

In order to generate an asymmetric structure of hexamers adjacent to pentamers in the lattice for structural characterisation of the peptide binding interfaces that face a pentamer, I modified the approach used to find pentamers in the lattice. Pentamers were detected as before using hexamer subtomogram positions, this time from alignment of the unbinned subtomograms, and these coordinates were used to calculate an orientation vector pointing towards the pentamer. An in-plane rotation corresponding to an integer multiple of 60° was then applied to each of the subtomogram positions based on the direction in which the orientation vector associated with it pointed, in order to place the pentamer at approximately the 12 o'clock position in the final map. The corresponding subtomograms from unbinned tomograms were then averaged using the revised set of Euler angles in order to generate the asymmetric references. A single iteration of asymmetric subtomogram alignment with parameters otherwise identical to the first hexamer alignment step on unbinned

data, as detailed previously, was performed using the asymmetric maps as references.

Curvature analysis: hexamer pair tilt and twist

The curvature of the CA lattice was analysed using the approach described by Mattei et al. (2016a) to determine the tilt and twist angles between all possible pairs of neighbouring hexamers. A bivariate histogram of tilt angle versus twist angle was then plotted and divided into discrete bins, which resulted in a matrix of structural classes which describe different combinations of local tilt and twist angles, corresponding to parts of the CA lattice with higher or lower curvature (Figure 4.6A). The subtomograms corresponding to each of the bins from the bivariate histogram of twist angles versus tilt angles were averaged to produce the corresponding classes. These classes describe curvature between the central hexamer in the map and one of its neighbours, but do not fully describe the local curvature of all the neighbours around the central hexamer.

In order to fully describe the local curvature of the capsid lattice around any given hexamer, I performed principal component analysis (PCA) in MATLAB on the tilt and twist angles of all neighbouring hexamers around any given hexamer. First, only hexamer subtomograms with a full complement of 6 hexamers surrounding them were chosen from the data set for which tilt and twist angles were determined. A $12 \times n$ matrix of tilt and twist angles corresponding to the neighbouring hexamers, in a fixed direction around the central hexamer, was then constructed, where n is the number of central hexamer positions to be considered. Eigendecomposition was then performed on this matrix to determine the most significant eigenvectors of this matrix, and the subtomogram positions with associated tilt and twist angles were projected onto the first two principal components to define new principal axes for the data. The corresponding PCA scores for each subtomogram position were plotted on a bivariate histogram of the first two principal component axes, and this histogram was divided into discrete bins to generate structural classes for averaging as before.

4.4 Results

4.4.1 *In vitro* CA assemblies with IP₆ are identical in structure to viral capsids

CA core morphology

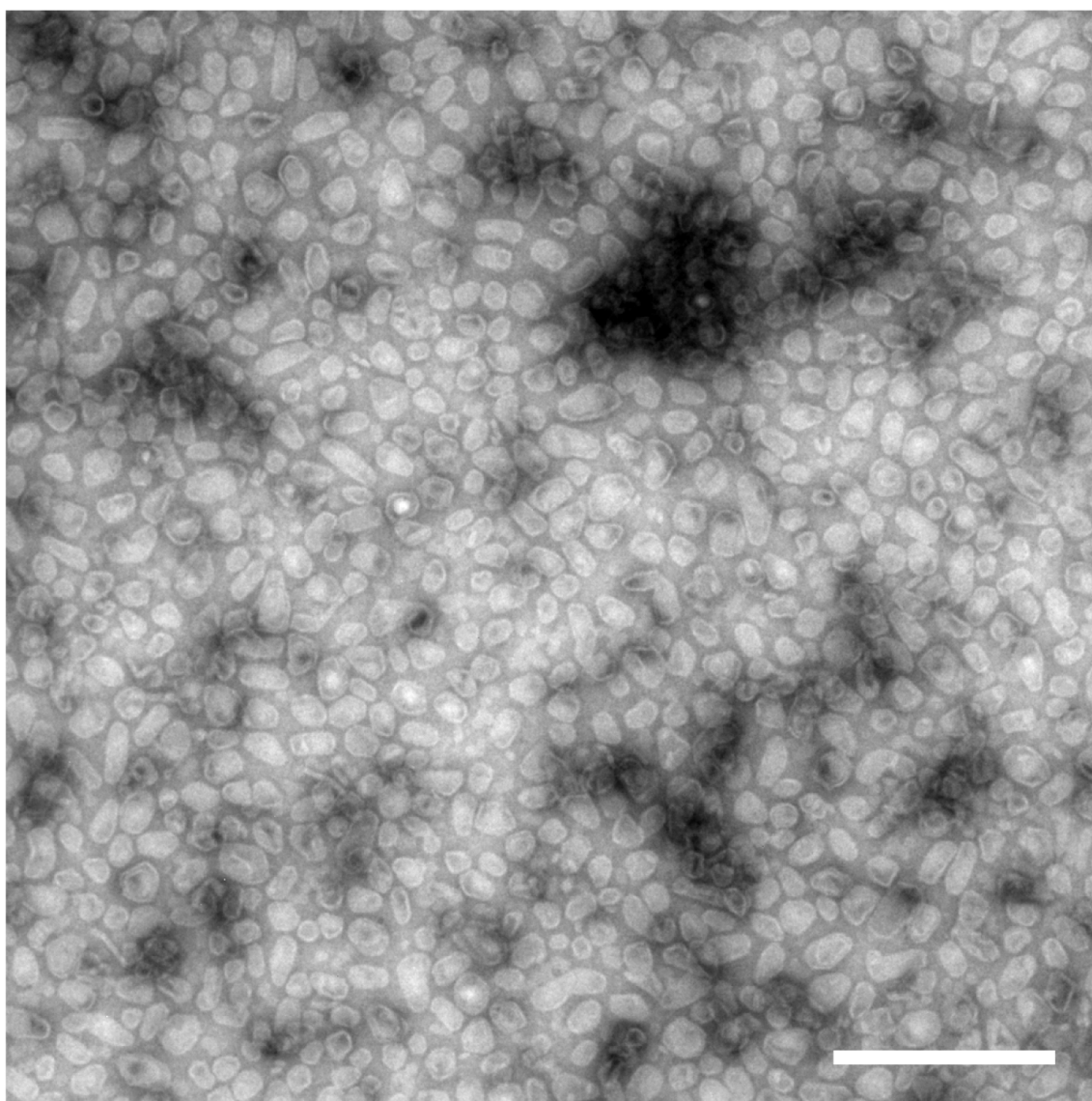


Figure 4.1: Representative negative stain EM image of *in vitro* CA cores assembled with IP₆. The overlaid scale bar corresponds to a distance of 500 nm in the image.

The morphology of the CA VLP cores was checked by negative stain electron microscopy prior to conducting cryo-EM experiments, and almost all cores were found to be conical in shape (Figure 4.1). The concentration of VLPs in the sample was also found to be very high, which agrees with the estimated CA protein concentration of 12 μ M. This

represents a significant improvement in throughput for structural determination compared with the relatively sparse distribution of capsids seen when using native viral preparations as previously done by Mattei et al. (2016a).

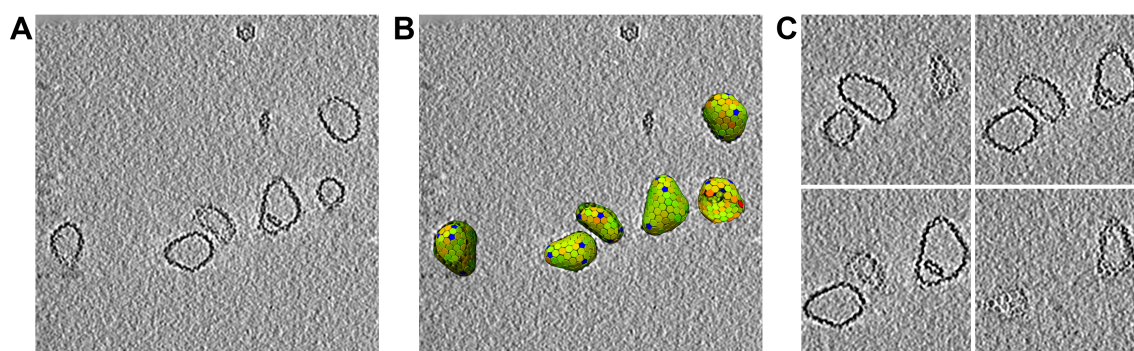


Figure 4.2: Computational slices through a tomogram of conical, *in vitro* CA cores. **A** A 1.1 nm thick computational slice through a representative tomogram, acquired at a target defocus of $-4.5\ \mu\text{m}$, showing the conical CA VLPs used for structural determination. The image here was rendered from an $8\times$ binned, non-CTF-corrected tomogram with a Gaussian filter applied (kernel width of 2 pixels) for display. **B** As in A, but with lattice maps overlaid to show the final aligned positions and orientations of the hexamers (red to green colour range, representing low to high cross-correlation coefficient) and pentamers (blue) found by reference-free subtomogram alignment. **C** 1.1 nm thick computational slices at different Z-height values through three of the VLPs in the same tomogram. Individual hexamers are clearly visible in the reconstructed tomograms, even before any subtomogram averaging.

Next, the morphology of the cores was assessed by cryo-ET and subtomogram averaging. Although the *in vitro*, IP_6 -stabilised conical CA assemblies here exhibit some variation in size and shape, they are overall comparable in dimensions to capsids in purified, authentic HIV-1 viruses (Mattei et al., 2016a). 5.6 %, 18.9 % and 14.3 % of the CA VLP cores with no peptide, 50 μM CPSF6 peptide or 50 μM Nup153 added, respectively, contained smaller cores nested within them. These nested VLPs were not used for structural determination as it is not known how this nesting of cores may affect accessibility of the cofactor binding sites on the inner core.

Examination of the aligned subtomogram positions and orientations as lattice maps (Figure 4.2) reveals that the morphology of the CA lattice found in these *in vitro* assemblies is also identical to that previously found in authentic viruses by subtomogram averaging (Mattei et al., 2016a). As with authentic HIV-1 viruses, the VLPs consist of a lattice of CA hexamers, closed by 12 pentamers in the case of a complete fullerene cone.

Mature CA hexamer structure in conical VLPs

I next applied subtomogram averaging to obtain a $3.9\ \text{\AA}$ structure of the mature hexamer from the conical CA cores assembled with IP_6 (Figure 4.3), a significant improvement in resolution over the in-virus structure previously obtained to $6.8\ \text{\AA}$ by Mattei et al. (2016a). This was made possible due to a number of reasons. First, the lack of a viral envelope,

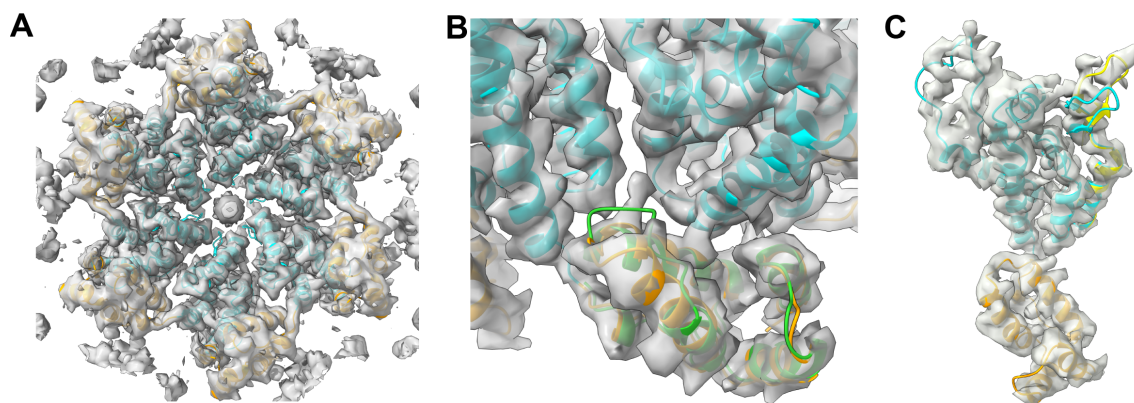


Figure 4.3: Mature CA hexamer structure at 3.9 Å resolution from IP₆-stabilised *in vitro* cores. **A** Isosurface view of the hexamer structure as seen from the CTD of CA. A backbone model of the mature hexamer (PDB accession code: 5MCY) (Mattei et al., 2016a) has been fit as a rigid body into the density (cyan: NTD, orange: CTD) and corresponds well to the map from the *in vitro* system. A strong central density is seen, which corresponds to IP₆. **B** Close up view of the density corresponding to the two CA molecules that form the binding pocket for CPSF6 and Nup153. The CTD of a crystal structure of CA with the binding pocket in the closed conformation (PDB accession code: 3H4E) is overlaid in green, showing that the density in this map corresponds to a pocket in the open conformation, which is able to bind CPSF6 and Nup153 (Price et al., 2014). **C** Zoned representation of the density showing the fit of a single CA protomer. The beta-hairpin density is in the open conformation in the *in vitro* cores, as shown by a superimposed atomic model of helix 1 and the beta-hairpin in the open conformation (yellow) (PDB accession code: 5HGK), which is the expected conformation at pH 6.0 (Jacques et al., 2016).

other viral proteins and RNP substantially improves the image quality obtainable using these cores, as having CA in isolation eliminates the image noise attributable to these other viral components. Another factor is the lower limit on ice thickness imposed by the dimensions of HIV-1 particles. The absence of a viral envelope in these VLPs means that they can be vitrified in a much thinner layer of ice than what is possible with purified viral samples, as HIV-1 particles are roughly spherical with an average diameter of between 120 and 130 nm, whereas the diameter of the capsid along its short axis is less than half of this, at around 57 nm (Nakai & Goto, 1996). The sample thickness at zero tilt is especially important for cryo-electron tomography, as the apparent thickness of the sample increases with tilt angle, leading to loss of contrast as explained in Chapter 1.

The third reason for the improvement in resolution is the combination of various improvements in image processing for subtomogram averaging compared to when Mattei et al. (2016a) first determined mature CA hexamer and pentamer structures from native viruses. These include the use of 3D CTF correction of the tomograms used, which has been shown to improve the resolution of immature CA structures (Turoňová et al., 2017), as well as better compensation for the missing wedge and sample geometry during subtomogram alignment using amplitude spectra of tomogram noise volumes (previously described in more detail in Chapter 3). A combination of all of these factors has therefore enabled this improvement in resolution.

The high-resolution map of the mature CA hexamer obtained corresponds well to the map

previously obtained by Mattei et al. (2016a) but reveals additional details in the form of clear amino acid side chain densities (Figure 4.3). Of particular interest is the binding pocket reported by Price et al. (2014) to interact with the nuclear import cofactors CPSF6 and Nup153, and which exists in both open and closed conformations. In the closed conformation, it was reported by Price et al. (2014) that residues 177-182, corresponding to the top of helix 8, unfold into a linker that obscures the binding pocket and prevents binding to CPSF6 and Nup153. The corresponding density in the hexamer structure obtained clearly corresponds to the open conformation, as shown by the fit of a crystal structure of the closed conformation into the map (Figure 4.3B).

The beta-hairpin domain has been suggested to gate access to a pore formed by a ring of six arginine residues (R18) that form the IP₆ binding site and may be involved in nucleotide import (Jacques et al., 2016). This domain exists in both open and closed conformations as shown by comparison of the various crystal structures in the literature, with the open conformation occurring below a pH of 7 and the closed conformation occurring above this pH Jacques et al. (2016). Examination of the hexamer structure that I obtained here shows that the beta-hairpin is in the open conformation (Figure 4.3C), which is consistent with the state predicted from the VLP assembly pH of 6.0.

Mature CA pentamer structure in conical VLPs

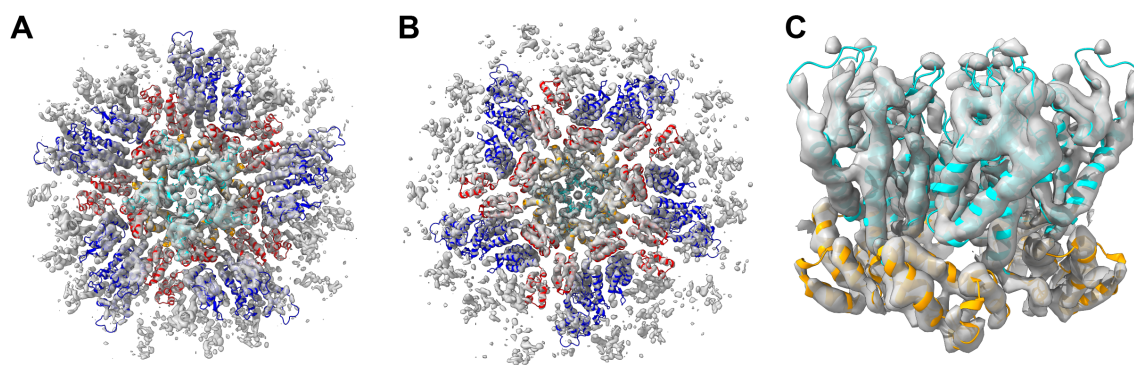


Figure 4.4: Mature CA pentamer structure at 6.2 Å resolution from IP₆-stabilised *in vitro* cores. **A** Iso-surface view of the pentamer structure as seen from the NTD of CA. A backbone model of the mature pentamer surrounded by 5 hexamers (PDB accession code: 5MCY) (Mattei et al., 2016a) is shown here, fit as a rigid body into the density (cyan: pentamer NTD, orange: pentamer CTD, blue: hexamer NTD, red: hexamer CTD). From this fit, it can be seen that the pentamer in the *in vitro* assembled cores is the viral pentamer, which is stabilised by inter-protomer interactions that are distinct from those seen in the crystal form of the pentamer (Mattei et al., 2016a). A strong central density is also visible in the structure, which may correspond to IP₆. **B** As in A, but viewed from the CTD of CA. **C** The density map zoned around only the pentamer, showing the fit of the backbone pentamer model obtained from authentic virions (as in A) into the density map from *in vitro* assembled cores with IP₆.

The pentamer structure obtained from the IP₆-stabilised CA VLPs was resolved to 6.2 Å (Figure 4.4), also a significant improvement over the viral pentamer structure previously

obtained by Mattei et al. (2016a) to 8.8 Å resolution. Crucially, however, the pentamer seen in these cores, which were assembled only from recombinant CA and IP₆, is the viral form of the pentamer, as opposed to the form obtained from previous attempts to crystallise the CA pentamer (Pornillos et al., 2011). As shown by Mattei et al. (2016a), helix 3 of the CA-NTD is not involved in the inter-protomer interface of the viral pentamer due to an approximately 19° rotation of each pentamer NTD relative to its position in the hexamer. This is in contrast to the crystal form of the pentamer, which exhibits quasi-equivalence with the hexamer and shares identical inter-protomer interfaces with it (Pornillos et al., 2011). The pentamer structure from these *in vitro* CA VLPs therefore shows that, given assembly conditions compatible with conical capsid assembly, an *in vitro* system consisting of CA and IP₆ is sufficient to result in the formation of viral pentamers.

Notably, these *in vitro* assemblies of CA lack the viral envelope that is normally an obstacle to performing capsid-binding experiments. This accessibility makes this *in vitro* system very useful as a model system to investigate interactions of host cofactors with not only hexamers, but also pentamers due to the biologically-relevant form of the pentamer being present here. In particular, the rotation of the CA-NTD in the viral pentamer results in large changes to the binding interface used by the host cofactors CPSF6 and Nup153, which raises important questions about whether pentamer-specific binding of these or other cofactors exists, and the effects such binding may have on capsid stability. This system is ideal as a tool to approach such questions, as the presence of only CA and IP₆ in the VLPs enables a bottom-up approach to investigating capsid binding by addition of individual cofactors in isolation.

4.4.2 Presence of IP₆ in the hexamer and pentamer structures

IP₆ is present in the CA hexamer across its range of curvatures

A strong density was seen in the central pore of the CA hexamer structure obtained from VLPs assembled in the presence of IP₆. This density was also present in the hexamers away from the six-fold symmetry axis in the structure, and in an asymmetric structure of hexamers next to a pentamer, which shows that it is not simply an artefact of the symmetrisation process (Figure 4.5B). The density in the central pore corresponds to that seen by Mattei et al. (2016a) in the hexamer from purified, authentic viruses in both size and shape.

Fitting a crystal structure (PDB accession code: 6BHT) into the density map shows that

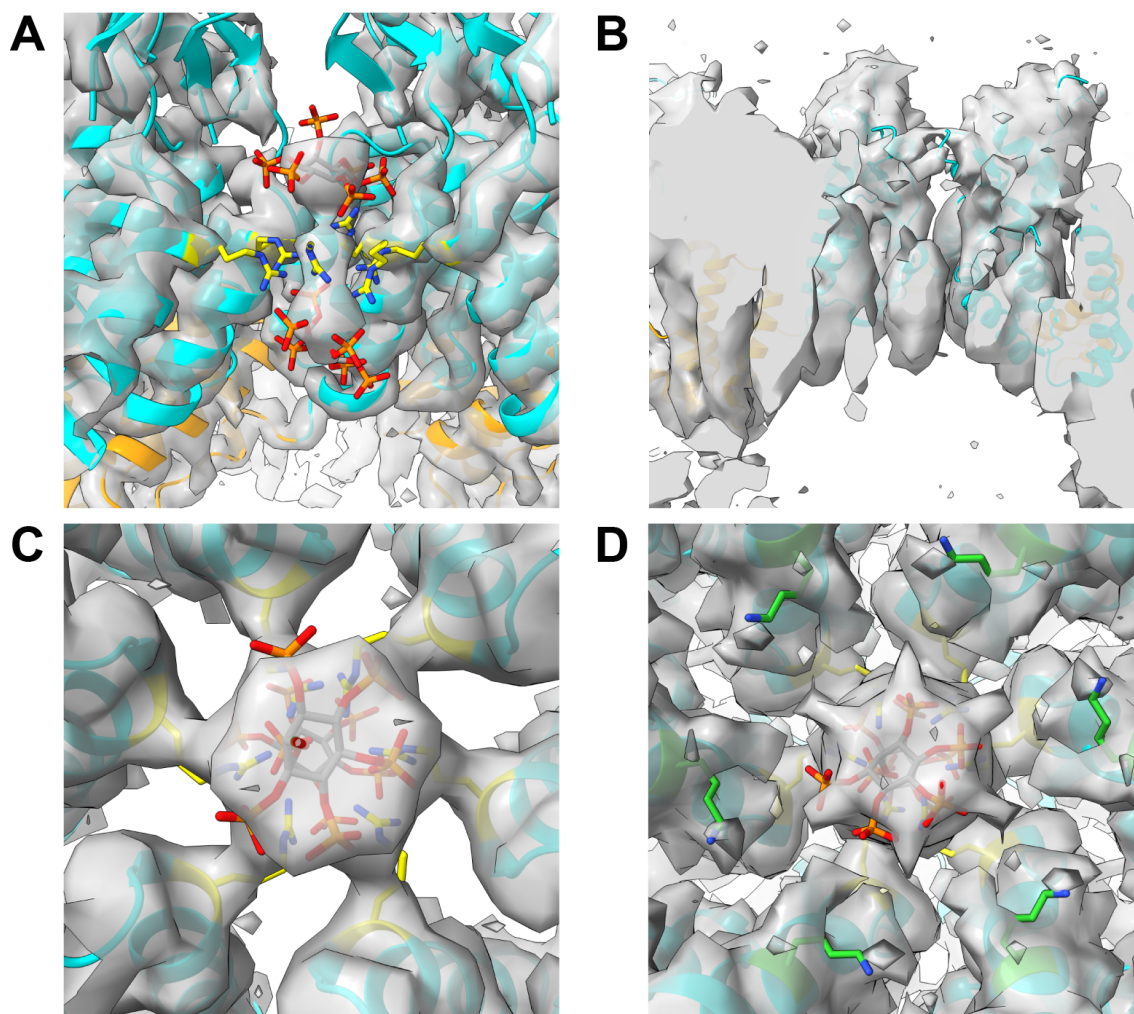


Figure 4.5: Comparison of the density in the R18 pore of the hexamer in VLPs with IP₆. **A** Density in the central pore below the beta hairpin in the CA hexamer structure shown in Figure 4.3, with a crystal structure of the CA hexamer bound to two molecules of IP₆ (PDB accession code: 6BHT) fit as a rigid body into the density map. The two IP₆ molecules from the crystal structure are shown in orange, with the arginine 18 side chains from the crystal structure depicted in yellow. The density in the hexamer pore from the structure obtained by subtomogram averaging is consistent with two IP₆ molecules being present in the pore. **B** Corresponding density within the central pore of a hexamer offset from the six-fold symmetry axis, showing that the central density is not a symmetry artefact. The density shown is from an intermediate step in the processing using $2\times$ binned subtomograms, before neighbouring hexamers were excluded from the alignment mask. **C** View of the map in A at a different isosurface threshold, showing the R18 side chains of the CA-NTD projecting in towards the IP₆ density. **D** View from the C-terminal end of the map in A, showing weak densities that appear at low threshold, which may correspond to a different rotamer of the K25 side chain from that shown in the model (green).

the density in the central pore of the hexamer is consistent with there being two IP₆ molecules, stabilised by electrostatic interactions with a ring of six arginine residues (R18) at the base of the beta-hairpin, and another ring of six lysine residues (K25) at the carboxy-terminal end of helix 1 (Figure 4.5C-D). Although the hexamer map is resolved to 3.9 Å, it is not possible to unambiguously determine the orientation of the IP₆ molecules in the six-fold symmetrised map. Another question raised by this is whether the occupancy of the IP₆ binding pore differs according to the local curvature, and thus

local structural flexing of the CA lattice.

In order to address these questions, I performed lattice curvature classification on the hexamer subtomograms as previously described by Mattei et al. (2016a) and extended it by performing PCA on the tilt and twist angles in order to describe all combinations of local lattice curvature in terms of two principal components, as described in Materials and Methods (Figure 4.6B). This has the effect of symmetry relaxation of the structure, which is important when trying to interpret the structure of the pore which is on the six-fold symmetry axis. Another advantage of taking this approach is that, unlike the curvature analysis described by Mattei et al. (2016a) in which the central hexamer and one neighbour are coherently averaged according to curvature, the relative orientations of all neighbouring hexamers around the central hexamer are considered in this PCA.

This curvature analysis was successful in classifying different parts of the CA lattice according to local curvature, but the main problem in interpretation of the data was the resulting resolution (Figure 4.6C). The resulting subdivision of the data set was not sufficiently offset by the improvement in B-factor from having more homogeneous classes, and the resolution obtained was therefore still insufficient to determine the exact mode of binding of IP₆. However, at the resolution of the maps achieved, there was no apparent change in the occupancy of the IP₆ binding pore regardless of lattice curvature, which suggests that IP₆ is uniformly bound to hexamers across the surface of the HIV-1 capsid.

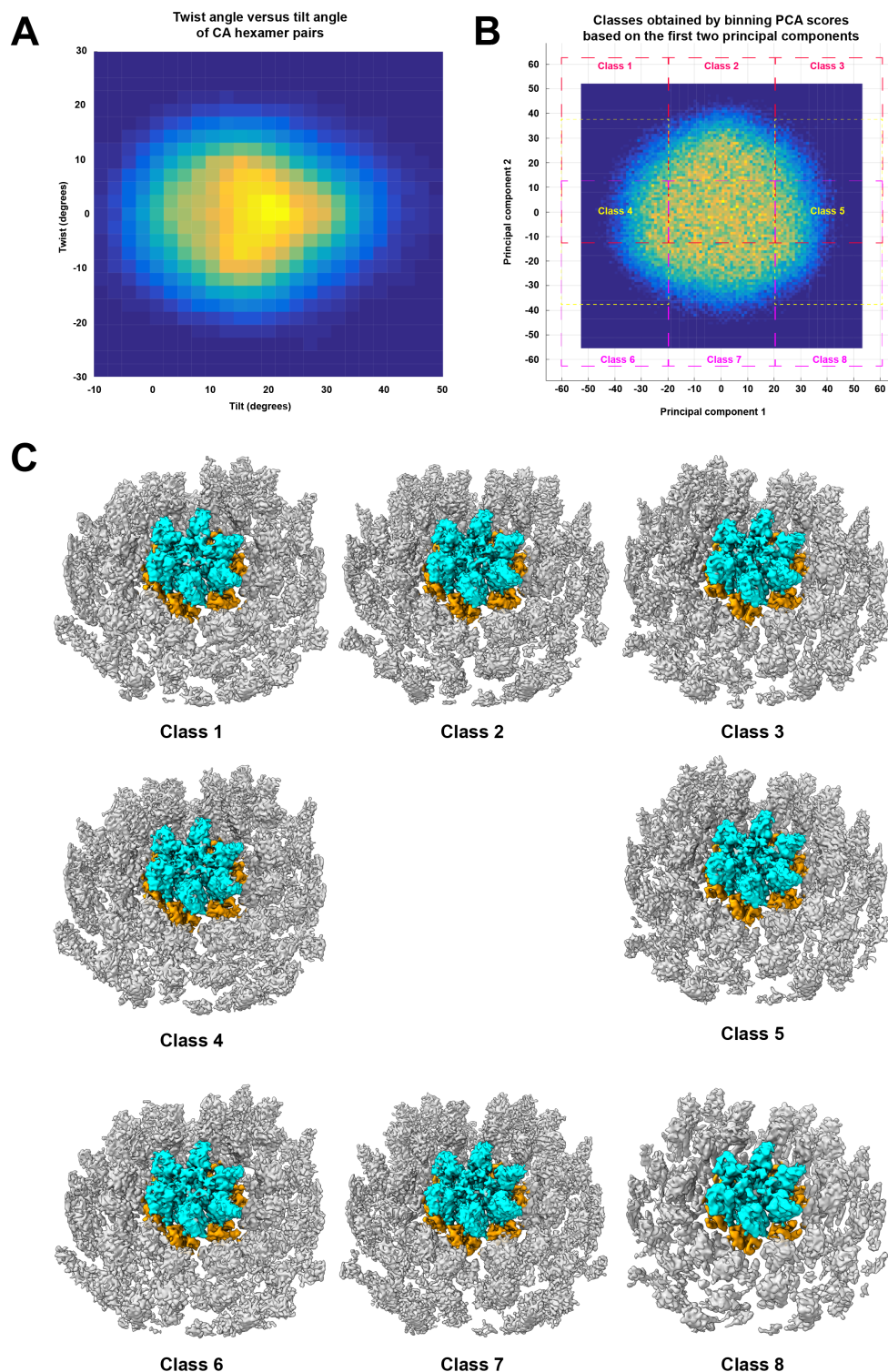


Figure 4.6: Lattice curvature classification of CA VLPs assembled with IP₆. **A** Bivariate histogram plotting different combinations of pairwise hexamer tilt and twist angles found by geometric classification of CA hexamers from VLPs assembled in the presence of IP₆. **B** Bivariate histogram plotting distribution of hexamer curvatures, obtained by using principal component analysis (PCA) to reduce the 6 pairs of neighbouring tilt and twist angles associated with each hexamer to a score corresponding to each of the first two principal components. Overlapping classes, denoted by dashed boxes of varying spacing, were obtained by grouping the subtomograms in different parts of the distribution. **C** Isosurface renders of the structural classes obtained by averaging the subtomograms grouped according to the distribution in B. The central hexamer is zoned according to domain, with the CA-NTD in cyan and the CA-CTD in orange.

HIV-1 CA pentamer structure exhibits density in IP₆ binding pocket

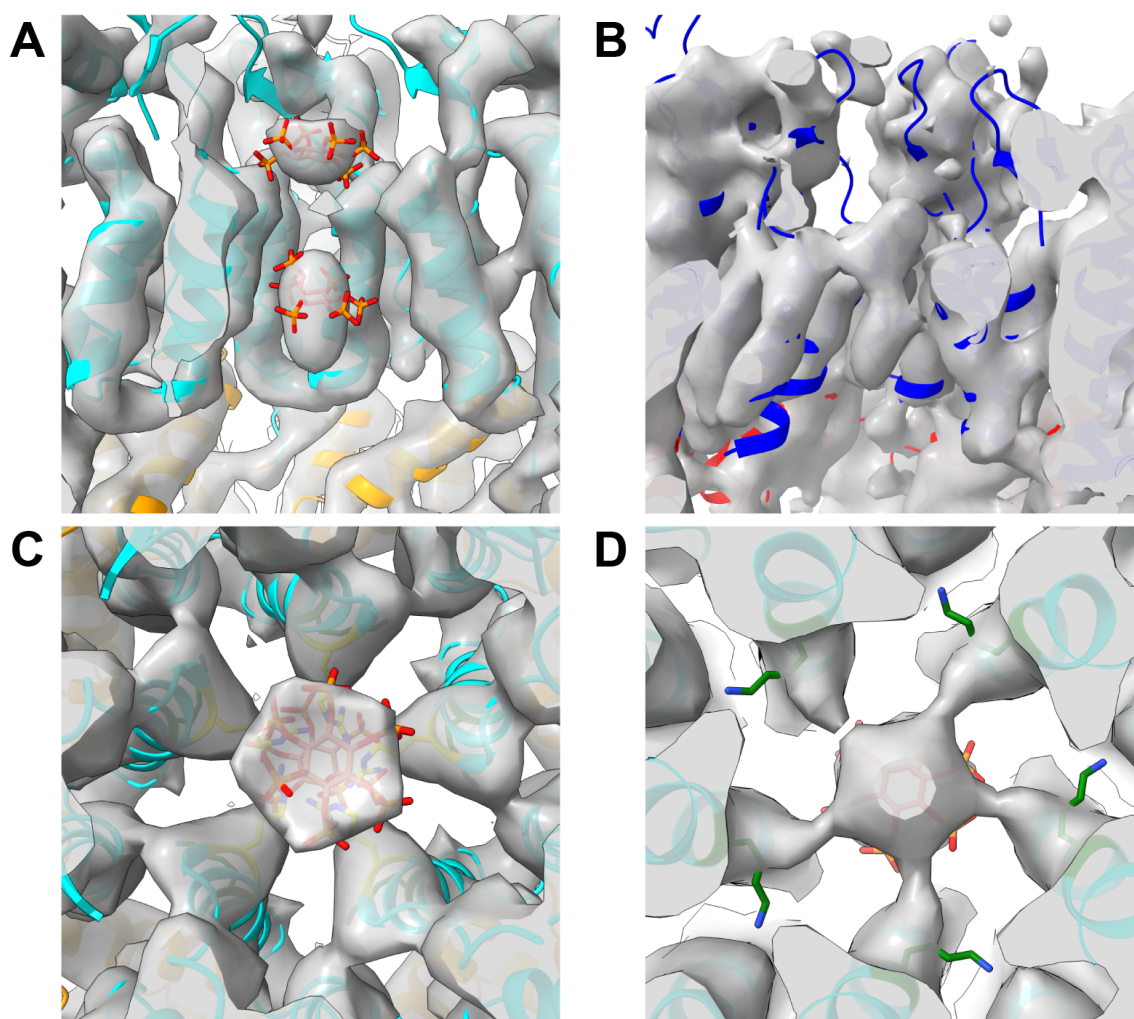


Figure 4.7: Comparison of the density in the R18 pore of the pentamer in VLPs with IP₆. **A** The CA pentamer map in Figure 4.4 also contains a strong central density in the equivalent R18 pore, similar to the hexamer from the same sample. The CA protomer from the crystal structure of the CA hexamer bound to two molecules of IP₆ (PDB accession code: 6BHT) was fit as a rigid body into the density map, with the two IP₆ molecules docked manually to show that the density in the pore is consistent with two molecules of IP₆ (shown in orange). **B** Corresponding density within the central pore of the pentamer resolved by centering pentamer-adjacent hexamers in the map. This density is offset from the five-fold symmetry axis that was used to generate the map in A, and shows that the density in A is not a symmetry artefact. **C** View of the map in A at a different isosurface threshold, showing density corresponding to the R18 side chains of the CA-NTD projecting in towards the IP₆ density. **D** View from the C-terminal end of the map in A, showing weak densities that appear at low threshold, which may correspond to a different rotamer of the K25 side chain from that shown in the model (green).

The inability to crystallise the pentamer in its viral form, and the low resolution obtainable to date due to the need to use purified viruses to obtain its structure by subtomogram averaging, have made it difficult to ascertain whether cofactors such as IP₆ also interact with the pentamer as they do with the hexamer. The use of an *in vitro* system that successfully reconstitutes the viral pentamer conformation within CA VLPs, stabilised by IP₆, allows this question to be addressed.

In the pentamer structure obtained, two strong densities are present within the central pore formed by the NTDs of the pentamer, as with the hexamer (Figure 4.7). These central densities are not symmetry artefacts, as they are still visible in an asymmetric structure of hexamers next to a pentamer, in which the pentamer is situated off-axis from the aligned asymmetric hexamer (Figure 4.7B). The identity of these densities in the pentamer pore cannot be conclusively determined, but they most likely correspond to two molecules of IP₆ and/or its degradation products, as there are no other polyanionic species in the core assembly buffer that would bind in such a way to the pore. This is supported by the presence of side-chain densities in the map that correspond to a ring of five arginine residues (R18) as well as a five lysine residues (K25), pointing towards the central pore densities and within the correct distance to form electrostatic interactions with them (Figure 4.7 C-D).

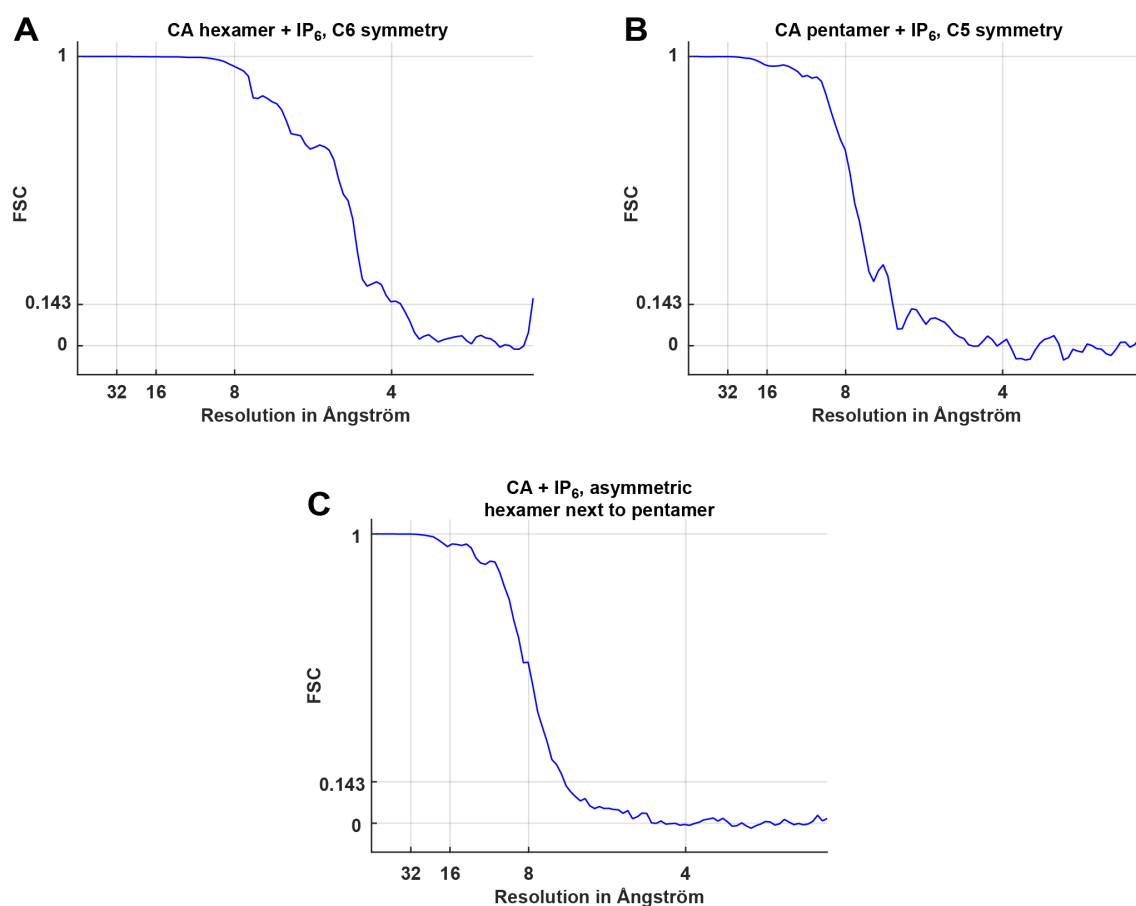


Figure 4.8: Fourier shell correlation (FSC) for structures obtained from CA VLPs with IP₆. **A** FSC for C6-symmetrized hexamer structure. **B** FSC for C5-symmetrized pentamer structure. **C** FSC for asymmetric structure of hexamers next to pentamers.

4.4.3 CPSF6 binds to CA hexamers but not pentamers

Addition of 50 μM CPSF6 peptide (residues 313-327) resulted in a hexamer structure with an extra density in the CPSF6 binding pocket which corresponds to the peptide (Figure 4.9B). This map was resolved to a resolution of 6.2 \AA , which is significantly lower than the hexamer map without any peptide bound (apo-hexamer), due to the smaller amount of data collected for the CPSF6-bound sample. However, at this resolution it is still clear that the mode of binding to the hexamer agrees well with a published crystal structure of *in vitro*, cross-linked CA hexamers with the same peptide added (PDB accession code: 4U0B).

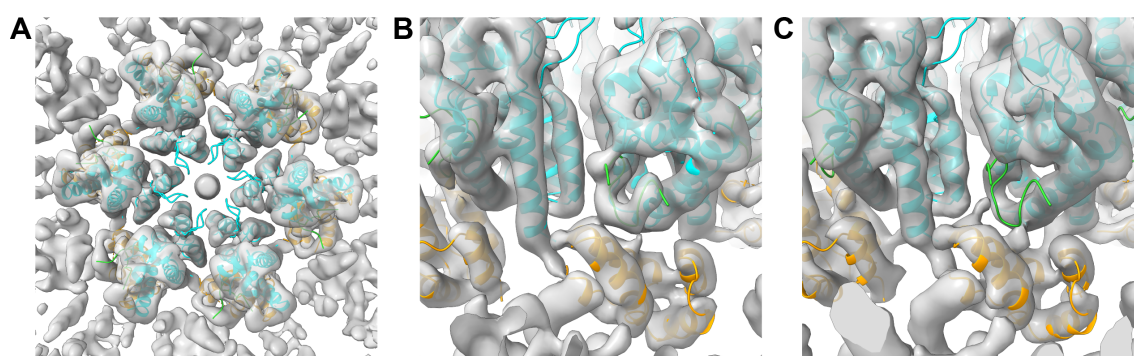


Figure 4.9: Structure of the CA hexamer from VLPs incubated with 50 μM CPSF6 peptide (residues 313-327). **A** Isosurface view of the density map of the CA hexamer incubated with a peptide consisting of the residues 313-327 of CPSF6, at 6.2 \AA resolution, as seen from the NTD of CA. Atomic models of the CA-NTD (cyan), CA-CTD (orange) and CPSF6 peptide (green) from a crystal structure of the CA hexamer bound to this peptide (PDB accession code 4U0B) were fit separately as rigid bodies into the density map. **B** Close up view of the binding pocket formed between one CA molecule and its neighbouring CA-NTD, showing the fit of the CPSF6 atomic model from A into the pocket. The density map obtained by cryo-ET and subtomogram averaging corresponds very well to the crystal structure. **C** The same atomic model from A and B shown superimposed on the density map of the hexamer from Figure 4.3 without any peptide added, low pass filtered to 6.2 \AA resolution, showing that the extra density in B corresponds to the CPSF6 peptide.

The CPSF6 peptide concentration of 50 μM used was chosen based on a recent finding that CPSF6 binds with over 10-fold higher affinity to whole HIV-1 capsids ($K_D = 3.4 \mu\text{M}$) than it does to individual hexamers previously used for equilibrium constant determination ($K_D = 50 \mu\text{M}$) (Márquez, 2019). This revised K_D value corresponds to an occupancy of approximately 92.4 % when the peptide is added at a concentration of 50 μM , assuming all binding sites are accessible. From the map obtained, it was not possible to quantitatively determine the actual occupancy of the CPSF6 peptide in the hexamer, but a qualitative analysis of the density in the peptide binding site suggests that, while present on average, the occupancy of the peptide is less than 100 %.

In contrast, no density corresponding to the CPSF6 peptide was found in the pentamer density map obtained from this sample (Figure 4.10A). However, the off-axis hexamers in the pentamer density map contain clear extra density corresponding to the CPSF6 pep-

tide (Figure 4.10B) at all sites facing the pentamer. The map quality at these sites was visibly worse than for the pentamer, as expected from their location close to the edge of the map volume, and therefore I obtained another structure by performing an alignment with pentamer-adjacent hexamers centred in the map. This was done by identifying the hexameric neighbours of a pentamer and then rotating each position by an appropriate multiple of 72° in order to orient the hexamer-to-pentamer vector towards the same direction in the final average. This structure shows that the CPSF6 peptide is present, on average, in the binding pocket facing the pentamer, but not in the structurally different binding pocket present in the pentamer itself (Figure 4.10C).

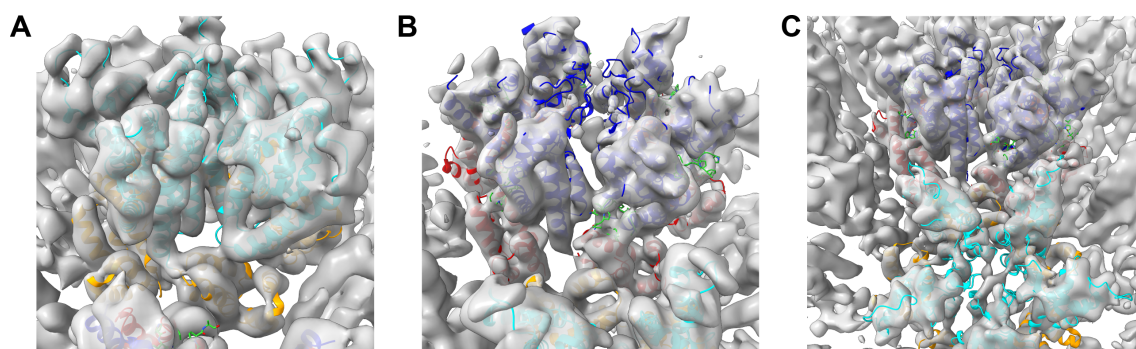


Figure 4.10: Binding pocket in the CA pentamer and adjacent hexamers with $50\mu\text{M}$ CPSF6 peptide (residues 313-327). **A** View of the binding pocket in the pentamer density map with $50\mu\text{M}$ CPSF6 peptide added, at 8.0 \AA resolution. There is no density in the map that corresponds to the CPSF6 peptide in the pentamer. **B** Off-axis, neighbouring hexamer in the pentamer structure shown in A. There is clear density in the map corresponding to the CPSF6 peptide in the binding pockets present in this hexamer. **C** 8.1 \AA structure obtained by centering the subtomogram alignment on hexamers next to pentamers, with the pentamer oriented in a single, defined direction. This asymmetric hexamer structure shows that the CPSF6 peptide is bound to all 6 sites on the hexamer next to a pentamer, with no evidence of differential binding across this range of curvatures.

In order to determine whether occupancy of CPSF6 in the hexamer changes with local lattice curvature, I performed a geometric classification of the CPSF6-bound hexamers based on tilt and twist angles between hexamer pairs. I then divided the bivariate histogram of twist angle versus tilt angle into discrete bins, and excluded the central peak from the analysis as this corresponds to the predominant combination of tilt and twist angles seen in the original, unclassified data (Figure 4.11B). I averaged the subtomograms corresponding to these bins in order to generate separate curvature classes, and examined the peptide binding site facing the neighbouring hexamer used for curvature determination, in order to investigate whether local lattice curvature can affect accessibility of the binding site.

In all of the classes obtained from this curvature analysis, the CPSF6 binding site in the central hexamer which directly faces the direction of lattice curvature was found to be in the open conformation, in which the N-terminal end of helix 8 remains folded (Márquez et al., 2018), and clear peptide density was visible in this site. No evidence of any change

in the mode of binding of CPSF6 based on lattice curvature was found.

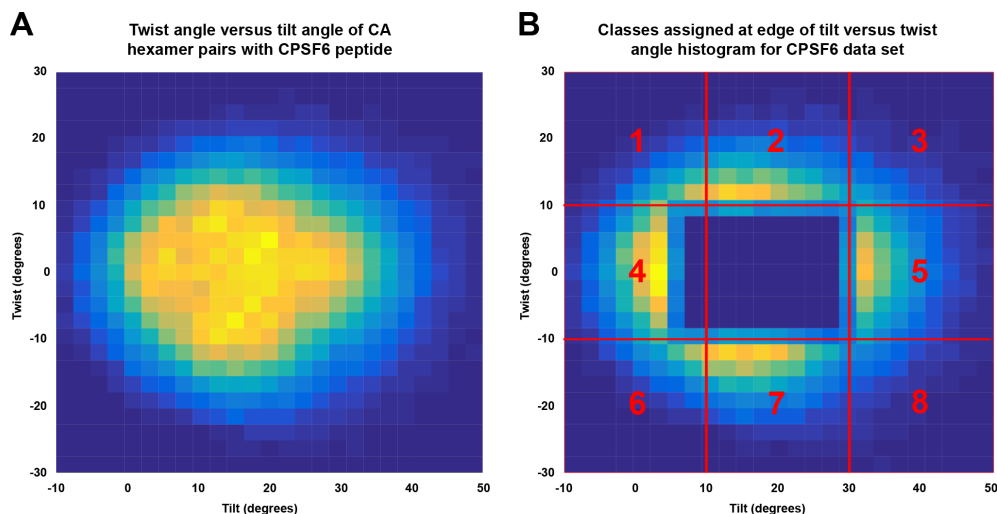


Figure 4.11: Bivariate histogram of hexamer pair twist angle versus tilt angle in CA VLPs with CPSF6 peptide. **A** Bivariate histogram plotting the frequency of different combinations of pairwise CA hexamer tilt and twist angles in the data set with CPSF6 peptide added. **B** As in A, but with the central peak removed, and the remaining subtomograms binned into 8 classes along the edges of the histogram (red grid with class numbers overlaid). Averaging these classes to produce structures with different lattice curvature revealed that the CPSF6 peptide density did not change with lattice curvature.

4.4.4 Nup153 also binds to hexamers but not pentamers

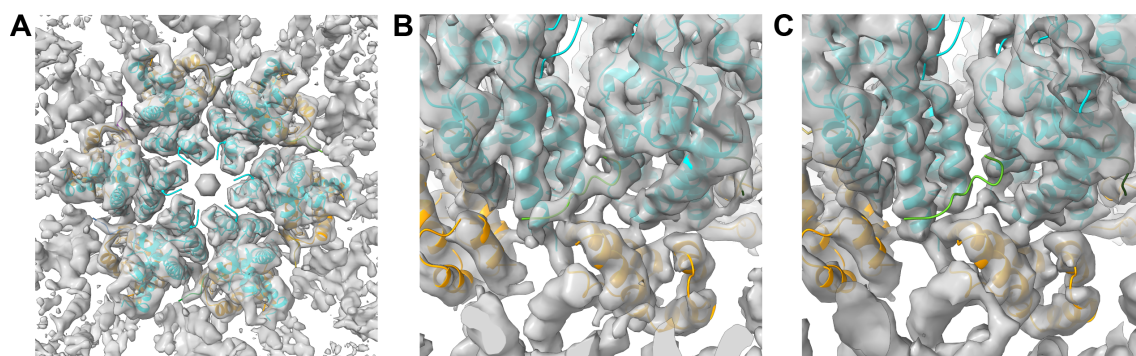


Figure 4.12: Structure of the CA hexamer from VLPs incubated with 50 μ M Nup153 peptide (residues 1407-1423). **A** Isosurface view of the density map of the CA hexamer incubated with a peptide consisting of the residues 1407-1423 of Nup153, at 4.8 Å resolution, as seen from the NTD of CA. Atomic models of the CA-NTD (cyan), CA-CTD (orange) and Nup153 peptide (green) from a crystal structure of the CA hexamer bound to this peptide (PDB accession code 4U0D) were fit separately as rigid bodies into the density map. **B** Close up view of the binding pocket formed between one CA molecule and its neighbouring CA-NTD, showing the fit of the Nup153 atomic model from A into the pocket. The density map obtained by cryo-ET and subtomogram averaging corresponds very well to the crystal structure. **C** The same atomic model from A and B shown superimposed on the density map of the hexamer from Figure 4.3 without any peptide added, low pass filtered to 4.8 Å resolution, showing that the extra density in B corresponds to the Nup153 peptide.

The resolution of the CA hexamer from VLPs bound to the Nup153 peptide (residues 1407-1423), at 4.8 Å, was considerably higher than that obtained with the CPSF6 peptide due to the larger size of the Nup153-bound data set, which was similar to that of the cores without peptide added. The structure of the hexamer clearly shows evidence of binding to the Nup153 peptide that is essentially identical to that seen in crystal structures (Figure 4.12B) (PDB accession code: 4U0D). In contrast to CPSF6, which binds mostly to a hydrophobic pocket in the N-terminal domain of one CA protomer and the neighbouring C-terminal domain, the Nup153 peptide extends further out of this pocket and forms more extensive contacts with the corresponding N-terminal domain of this neighbouring CA protomer (Figures 4.9B and 4.12B) (Price et al., 2014).

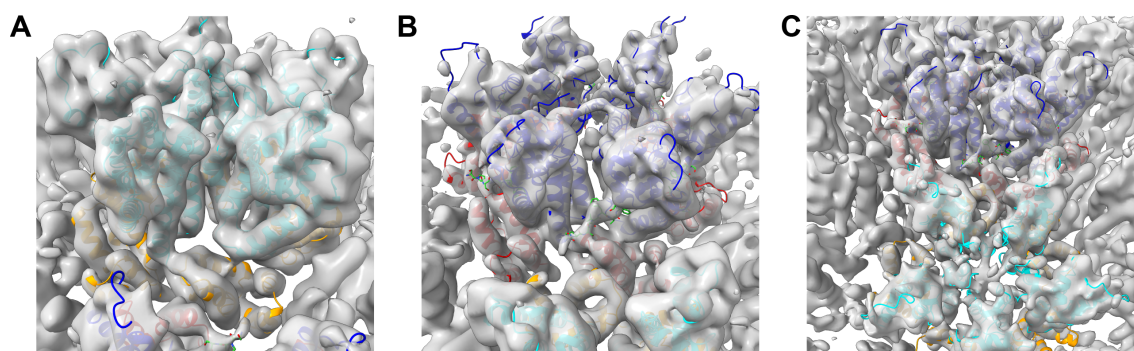


Figure 4.13: Binding pocket in the CA pentamer and adjacent hexamers with 50 μ M Nup153 peptide (residues 1407-1423). **A** View of the binding pocket in the pentamer density map with 50 μ M Nup153 peptide added, at 7.0 Å resolution. There is no density in the map that corresponds to the Nup153 peptide in the pentamer. **B** Off-axis, neighbouring hexamer in the pentamer structure shown in A. There is clear density in the map corresponding to the Nup153 peptide in the binding pockets present in this hexamer. **C** 7.2 Å structure obtained by centering the subtomogram alignment on hexamers next to pentamers, with the pentamer oriented in a single, defined direction. This asymmetric hexamer structure shows that the Nup153 peptide is bound to all 6 sites on the hexamer next to a pentamer.

Similar to the pentamer structure from VLPs with CPSF6 peptide, there was also no detectable peptide density in the pentamer density map from CA cores bound to Nup153 peptide at 50 μ M (Figure 4.13A). However, Nup153 peptide binding was also seen in all of the visible binding sites of the hexamers surrounding pentamers (Figure 4.13B). The presence of the peptide in the binding pocket facing the pentamer, but absence from the pentamer itself, was confirmed from the structure obtained by orienting and centering pentamer-adjacent hexamers followed by alignment on the re-centred map (Figure 4.13C), as described previously for the CPSF6-bound VLPs.

Nup153 hexamer binding changes with curvature

As with the CPSF6 data set, I performed a curvature classification on the Nup153-bound CA hexamer subtomogram coordinates according to pairwise tilt and twist angle combinations (Figure 4.14B). This yielded structural classes sorted according to the same

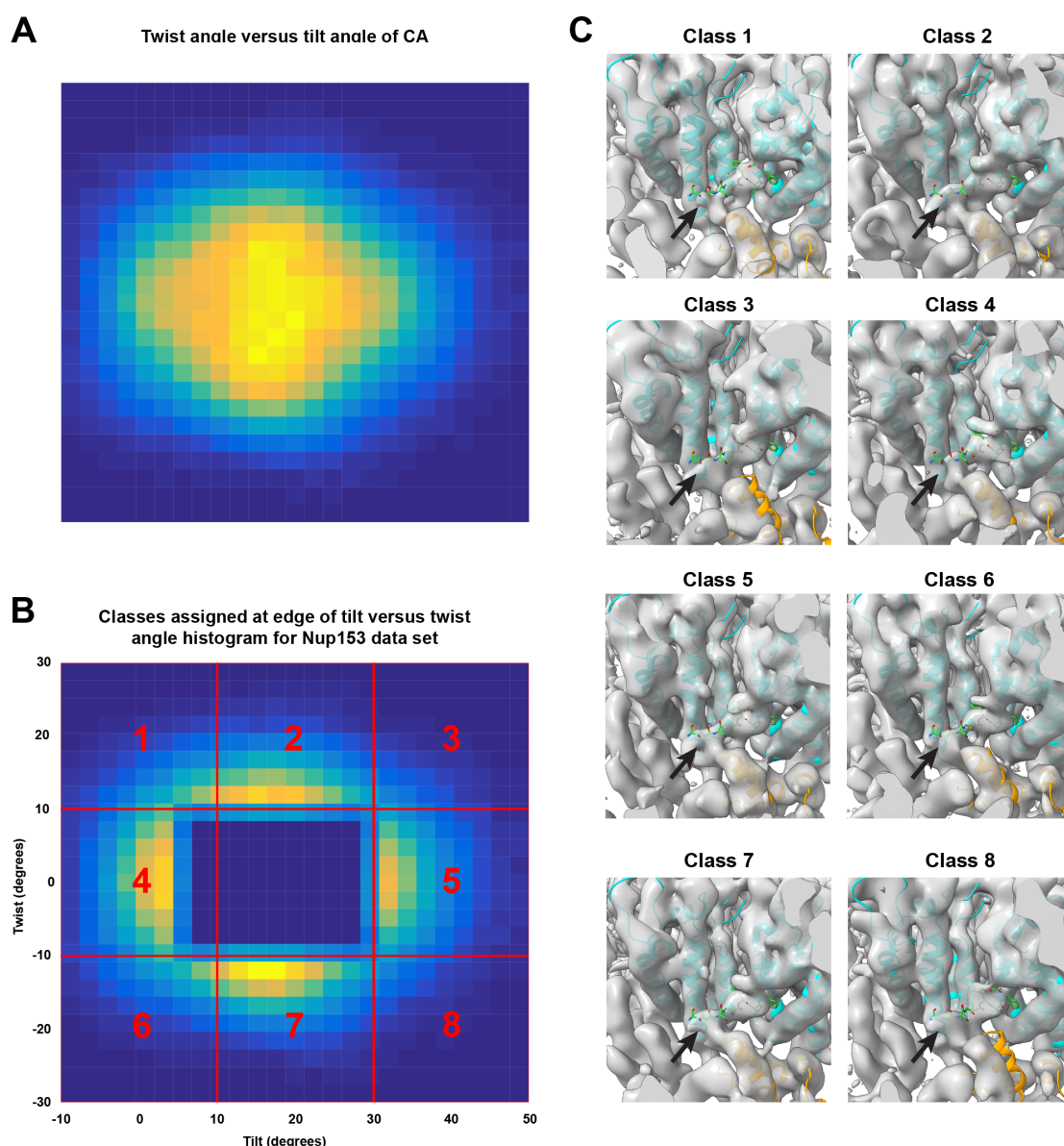


Figure 4.14: Bivariate histogram of hexamer pair twist angle versus tilt angle in CA VLPs with Nup153 peptide. **A** Bivariate histogram plotting the frequency of different combinations of pairwise CA hexamer tilt and twist angles in the data set with Nup153 peptide added. **B** As in A, but with the central peak removed, and the remaining subtomograms binned into 8 classes along the edges of the histogram (red grid with class numbers overlaid). **C** Classes generated by averaging the subtomograms grouped according to the classes shown in B. The NTD (cyan) and CTD (orange) from a crystal structure of CA bound to the same Nup153 peptide (green) (PDB accession code: 4U0D) are shown, fit separately as rigid bodies into the density map corresponding to each class. The Nup153 peptide coordinates were fit as a single unit together with the CA-NTD in order to prevent spurious fits into the density map, due to its small size. The length of the density corresponding to the peptide can be seen to vary in length when the maps are thresholded to show the same level of detail in the surrounding protein regions (black arrow overlaid in the same place in each map for comparison).

curvature criteria previously used for CPSF6. As with CPSF6, every class generated contained the binding pocket in the open conformation, in which the top of helix 8 is folded and does not occlude the binding site (Figure 4.14C).

As previously reported from a crystal structure of the same Nup153 peptide sequence

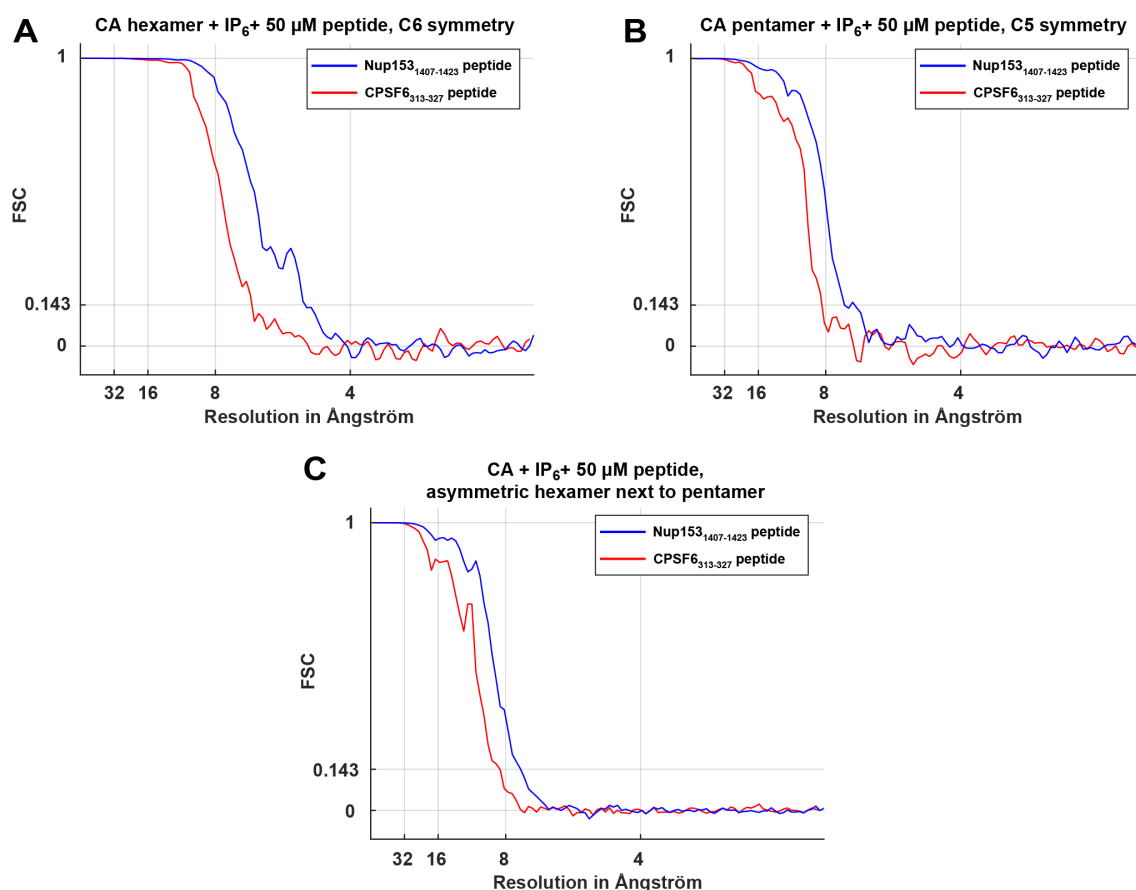


Figure 4.15: Fourier shell correlation (FSC) for CA structures with IP₆ and 50 μM Nup153 (blue) or CPSF6 peptide (red) added to the CA VLPs. **A** FSC for C6-symmetrised hexamer structure. **B** FSC for C5-symmetrised pentamer structure. **C** FSC for asymmetric structure of hexamers next to pentamers.

bound to cross-linked CA hexamers, Nup153 interacts with many CA residues in common with CPSF6 (Price et al., 2014). This includes the main, critical binding site to the CA protomer, which is a pocket formed by helices 3, 4 and 5 as well as the turn between helices 5 and 6 in one CA-NTD, with which both F1417 of Nup153 and F321 of CPSF6 form extensive hydrophobic contacts (Price et al., 2012; Price et al., 2014). However, CPSF6 has an almost cyclic structure and the other interactions outside of this hydrophobic pocket are mainly formed with the CTD of the neighbouring CA molecule, whereas the extended structure of the Nup153 peptide mainly interacts with residues on the neighbouring NTD instead (Price et al., 2014). The curvature analysis that I performed shows that at some combinations of tilt and twist angles, the part of the peptide density outside of the hydrophobic pocket, corresponding to Nup153 residues 1410 to 1413 which bind mainly with residues in the NTD of the neighbouring CA molecule, weakens considerably (Figure 4.14C). This effect is visible when two neighbouring hexamers lie almost flat with tilt angles between -10° and 10° , and when a hexamer pair has a tilt of between 10° and 30° but a strong twist of between -30° and -10° .

Although curvature-based classification of the data set allowed these classes to be re-

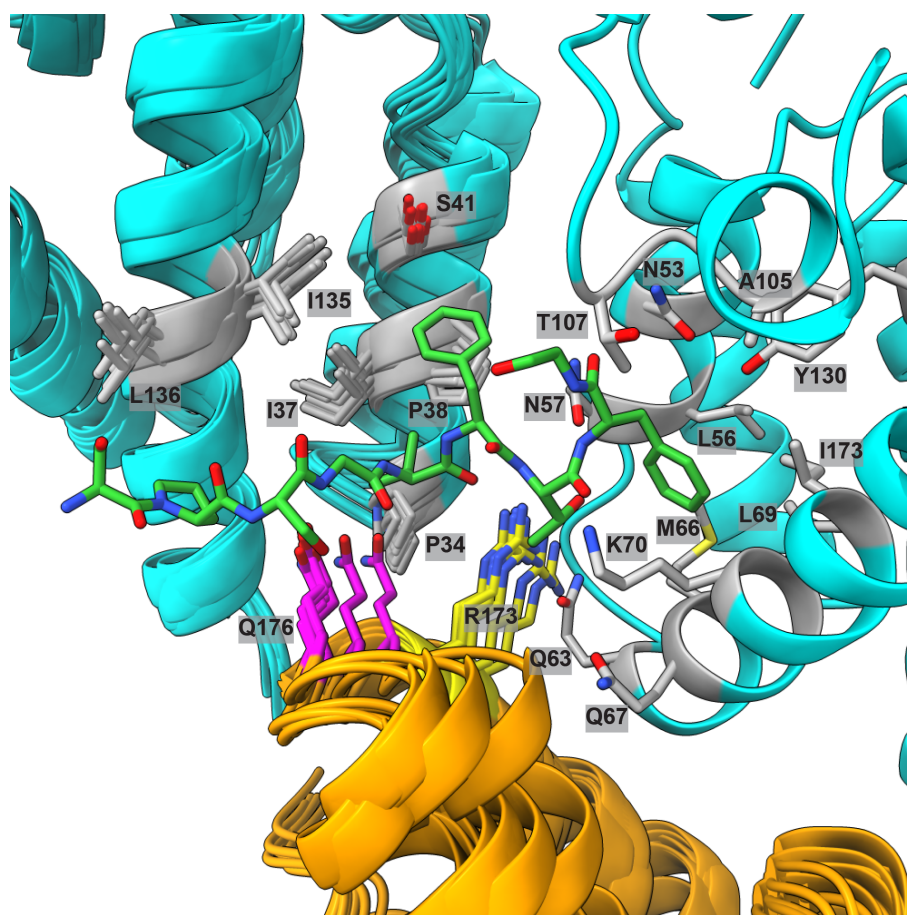


Figure 4.16: Variability in the side chain positions of the neighbouring CA molecule with curvature in the Nup153 binding site. The separately fitted NTD (cyan) and CTD (orange) positions, relative to the right-most NTD in the image, are shown for the different curvature classes in Figure 4.14C. The side chains shown correspond to those identified by Mallery et al. (2018) by crystallography to interact with Nup153. R173 (yellow) and Q176 (magenta) are highlighted to show the variability in the positions of these specific residues as lattice curvature changes.

solved, the resulting subdivision of the subtomograms across multiple classes also resulted in a decrease in resolution. It was therefore not possible to conclusively state how the mode of binding of the Nup153 peptide changes across the range of curvatures based directly on amino acid side chain densities in the maps. However, as with the CPSF6 curvature classification, the maps were of sufficient quality to perform rigid body fitting of the individual CA-NTDs and CTDs that comprise the binding interface in each map, and to then compare the relative movement of these domains across the curvature classes. Based on this analysis, the positions of residues R173 and Q176 in helix 8 of CA were quite variable (Figure 4.16). It is known from crystal structures that both of these residues form interactions with the Nup153 peptide (Price et al., 2014), pointing to a possible role of this range of movement in destabilising binding of CA to part of the peptide. The intensity of the side chain density corresponding to R173 in the maps was also variable, but no strong correlation between this and the occupancy of Nup153 residues 1410-1413 could be deduced. Crucially, however, this analysis did not detect any local lattice curva-

ture at which Nup153 binding was completely abolished, with only the part of the peptide outside the critical hydrophobic pocket affected by side chain movement.

4.5 Discussion

4.5.1 Modulation of capsid stability

The stability of the HIV-1 capsid is of critical importance to the viral life cycle and successful infection. This has been shown by experiments using fluorescently-tagged CypA and integrase to track capsid disassembly and nuclear entry in single viruses, providing evidence that successful nuclear import and infection are associated with a significant amount of intact CA lattice at the nuclear pore and that premature uncoating leads to viral degradation by proteosomes (Francis & Melikyan, 2018). On the other hand, point mutations in the CA protein that result in hyperstable capsids with increased stiffness and altered uncoating rates are clearly associated with a significant reduction in viral infectivity (Ambrose et al., 2012; Forshey et al., 2002; Ramalho et al., 2016). The timing of capsid uncoating is therefore crucial if the virus is to successfully achieve nuclear entry and integration of its genome into host chromatin.

Capsid stability is modulated by a variety of different cofactors present in the cellular environment. IP₆ is one such cofactor that has been of great interest in recent years, as this abundant cellular polyanion is not only an important assembly cofactor for immature viral assembly (Dick et al., 2018), but also dramatically increases the stability of the capsid and has been proposed to act in a manner similar to picornavirus pocket factors by influencing capsid stability and uncoating (Mallery et al., 2018). Crystal structures of the mature CA hexamer with IP₆ show that it neutralises the positive charge concentrated in a pore formed by six arginine residues in CA (Dick et al., 2018; Mallery et al., 2018). However, it is unknown whether IP₆ interacts similarly with the pentamer, as crystallisation of the form of the pentamer present in the virus has proven difficult (Pornillos et al., 2011), and structural determination of the viral pentamer by subtomogram averaging has been limited in resolution (Mattei et al., 2016a).

Here, I have demonstrated the use of an *in vitro* system of conical CA assemblies containing IP₆, which significantly improves the throughput and resolution of CA hexamer and pentamer structural determination by subtomogram averaging. The advantage of this system is that the CA lattices formed faithfully replicate the variability in curvature seen in native capsids and also incorporate the viral form of the pentamer (Mattei et al., 2016a). Using these CA VLPs, I have obtained the structure of the hexamer to a resolution of 3.9 Å

and the structure of the pentamer to a resolution of 6.2 Å. The hexamer structure contains density in the central R18 pore that is consistent with the presence of two molecules of IP₆, in agreement with the P2₁2₁2₁ crystal form of the CA hexamer which contains one molecule of IP₆ above the R18 ring and another below it (Figure 4.5) (Dick et al., 2018). Crystal structures with IP₆ bound only to the site above the R18 ring have also been reported (Dick et al., 2018; Mallery et al., 2018), and it is possible that the high occupancy of IP₆ in both positions in the hexamer structure from these VLPs is due to the very high concentration of IP₆ added during assembly (2.5 µM). I also determined high-resolution structures of the hexamer at various lattice curvatures using this VLP system, which show that IP₆ binding in the mature hexamer does not appear to be affected by the range of movements in the lattice needed to accommodate its natural range of curvatures.

The structure of the pentamer obtained in this manner also contains two equivalent densities in the same configuration and the ring of R18 side chains can also be seen extending towards the densities (Figure 4.7). This is consistent with the central density seen in a previously published, lower resolution structure of the pentamer obtained from authentic viruses (Mattei et al., 2016a). As with the hexamer, these densities were positioned above and below the ring of R18 side chains, and no bulky side chains which could have explained this density were found to have rotated into the pentamer pore due to the 19° rotation of the CA protomers between the hexamer and the pentamer. As no other negatively charged polyanions were added to the assembly reaction, it is very probable that the density in the R18 pore of the pentamer corresponds to two molecules of IP₆, or possibly a degradation product thereof such as inositol pentakisphosphate (IP₅). Due to the resolution of the map, it is not possible to conclusively ascertain the mode of binding of IP₆ to the pentamer. However, given that the pore in the pentamer is large enough to accommodate either the planar or sideways IP₆ orientations seen in the hexamer (Figure 4.7) (Dick et al., 2018), its mode of binding is unlikely to be very different from that in the hexamer.

4.5.2 Nuclear import cofactors Nup153 and CPSF6

The presence of only CA protein and IP₆ in the assembly system also enables these VLPs to be used in a bottom-up approach to structural characterisation of cofactors that bind to the capsid and potentially influence its stability as well as the nuclear import process. CPSF6 and Nup153 are host cofactors that interact with the capsid, potentially modulating its stability as part of their role in nuclear import. By incubating IP₆-stabilised CA VLPs with peptides containing the CA-binding sequences of CPSF6 and Nup153, I have obtained high-resolution structures of the CA hexamer and pentamer in cores that struc-

turally mimic authentic capsids. This provides direct evidence that, at a concentration corresponding to approximately 92.4 % CPSF6 or Nup153 binding site occupancy in the CA hexamer, neither of these peptide sequences visibly interact with the form of the CA pentamer found in authentic viruses (Figures 4.10 and 4.13). This finding is not entirely unsurprising, given that the CA protomers within the pentamer are rotated relative to their positions in the hexamer, resulting in solvent exposure of the hydrophobic pocket in the CA-NTD that is a key part of the binding site for both of these cofactors (Mattei et al., 2016a; Price et al., 2014). However, this result does have important implications for our understanding of HIV-1 nuclear import, as it is the first direct evidence that the relevant binding sequences from CPSF6 and Nup153 do not bind to the pentamer at concentrations at which consistent hexamer binding is achieved.

It is tempting to speculate that the more exposed CPSF6/Nup153 binding interface present in the pentamer may play a role in triggering uncoating once the capsid has been trafficked to the nuclear pore complex, via specific recognition of pentamers at the narrow tip or broad base of the capsid. However, this does not seem to be the case for Nup153 and CPSF6, as the complete lack of binding of the hexamer-binding sequences of these cofactors to the pentamer does not support such a hypothesis. This does not, however, rule out the possibility that other, as of yet unknown cofactors may specifically interact with pentamers and potentially influence lattice stability, or that a different part of CPSF6 and/or Nup153 may interact with the pentamer. In order to test this, it is important that a method to assemble viral pentamers *in vitro*, outside the context of the surrounding hexameric lattice, is developed in order to facilitate pull-down and binding affinity experiments.

It is also possible that nuclear import cofactors interact differently based on the local curvature of the lattice, which has been shown by Mattei et al. (2016a) to be accommodated by a combination of small movements of the CA-NTD and CTD relative to each other about the flexible linker that separates them. This may result in changes to the interfaces involved in binding CPSF6 and Nup153, especially when these movements become larger with more extreme lattice curvature. The use of these conical CA VLPs has enabled the binding of the CPSF6 and Nup153 peptides to distinct local curvature states of the capsid lattice to be characterised by applying geometric subtomogram classification. In this way, I have shown that there appears to be no change in the binding occupancy of the CPSF6 peptide used across different hexamer curvatures present in the lattice.

In contrast, the Nup153 peptide used does appear to have a lattice curvature-dependent mode of binding, with relatively flat regions of the lattice (hexamer-hexamer tilt angles between -10° and 10°) exhibiting a considerable weakening of the peptide density furthest from the hydrophobic pocket shared by CPSF6 and Nup153 (Figure 4.14). Based on both these cryo-EM structures as well as previously published crystal structures of the CA

hexamer in complex with these peptides (Price et al., 2014), this result is consistent with the CPSF6 peptide adopting a more cyclic conformation than Nup153, which is more linear and projects further out of the hydrophobic binding pocket towards the NTD of the neighbouring protomer. The weaker region of density corresponds to the part of the Nup153 peptide which interacts with the neighbouring CA protomer, mostly in the NTD but also with a number of CTD interactions (Price et al., 2014). This suggests that the binding occupancy of Nup153 is lower along the flat edges of the capsid compared to the more curved regions such as the tip and base. It can also be speculated that subtle changes in binding occupancy of factors such as Nup153 may facilitate exchange of these factors during different stages of nuclear import, as different cofactors are responsible for mediating different steps in this pathway. The FG sequence that binds to this pocket is a very common feature of nucleoporins, with various FG-nucleoporins found in the central channel, cytoplasmic side and basket of the nuclear pore (Aramburu & Lemke, 2017). Hence, although changes in binding occupancy may be subtle for any given cofactor, the abundance of the binding sequence at the nuclear pore may lead to competition that can result in different cofactors preferentially binding to different parts of the capsid, as the data here indicate may be the case. Disrupting this balance could therefore be a potential target for therapeutic intervention.

4.5.3 Implications for therapeutic design

The CPSF6 and Nup153 binding interface in the mature HIV-1 CA hexamer is of great interest as a target for antiretroviral therapeutics. One compound in particular, PF74, contains the same FG motif used by CPSF6 and Nup153 to bind to the hydrophobic pocket in the CA-NTD (Bhattacharya et al., 2014). PF74 has a bimodal effect depending on its concentration, inhibiting reverse transcription at low micromolar concentrations (Blair et al., 2010) but also affecting the mechanical stability of the capsid at concentrations of 10 μ M and above (Shi et al., 2011). Single-capsid uncoating assays have suggested that PF74 induces formation of defects in the capsid, but after initial defect formation, the remaining capsid lattice is greatly stabilised (Márquez et al., 2018).

Similar to CPSF6 and Nup153, PF74 binds across two protomers within a CA hexamer, and one consequence of this is that it has 22-fold higher affinity for the CA hexamer compared to monomeric CA protein (Price et al., 2014), which stands in contrast to similar compounds such as BI-2 which do not bind across intra-hexamer CA interfaces and thus do not selectively bind to assembled capsids. The exact mechanism by which PF74 induces capsid defects while stabilising the remaining lattice is still unknown, but Márquez et al. (2018) speculated that PF74 may stabilise certain strained conformations incompat-

ible with regions of high curvature, while the stabilising effect may be due to how PF74 bridges CA protomers within a hexamer. A possible mode of action compatible with this hypothesis is selective interaction of PF74 with pentamers, which are naturally incorporated into lattice regions of high curvature and would be an obvious target, but it has not been possible to obtain structural evidence to confirm whether this is true.

The utility of the IP₆-stabilised CA VLPs presented in this chapter has been demonstrated in investigating whether Nup153 and CPSF6-derived peptides interact with the viral form of the CA pentamer, which has a much more exposed binding pocket for these cofactors than in the hexamer. Although no binding was detected with these peptides, PF74 has been shown by isothermal calorimetry to bind to the hexamer with a K_D that is over $400\times$ lower than that of CPSF6 (Márquez, 2019; Price et al., 2014), and this much higher affinity may reveal potential pentamer interactions if the problem with the CPSF6 and Nup153 experiments was simply one of reduced affinity due to the more exposed binding site in the pentamer. I have also demonstrated that this VLP system allows structural characterisation of differential cofactor binding to the capsid based on lattice curvature. This system is therefore ideal as a way to investigate curvature-dependent effects as well as the possibility of pentamer binding with PF74 and its derivatives. Such experiments would be of great use in guiding structure-based drug design, as most of the current high-resolution structural information on capsid cofactors in the literature comes from X-ray crystallography experiments, which do not adequately account for capsid curvature effects despite increasing evidence that these may play an important role in successful HIV-1 infection.

Conclusions and perspectives

The work in this thesis sheds light on mechanistic aspects of immature HIV-1 assembly, maturation initiation, stability of the mature capsid and its interaction with host cofactors to modulate capsid stability and nuclear import. The chapters in this thesis have been presented in chronological order according to the work I performed.

In Chapter 2, I applied improvements in cryo-ET data collection and processing methods to determine high-resolution structures of a panel of Gag cleavage mutant constructs previously only characterised to low resolution by de Marco et al. (2010). This has helped to elucidate the link between proteolytic and structural maturation, and showed that CA-SP1 helical bundle destabilisation is the structural switch for maturation.

Chapter 3 presents classification experiments I performed on a published cryo-ET data set of wild-type immature HIV-1 viruses (Schur et al., 2016). To do this, I implemented the WMD PCA classification algorithm (Heumann et al., 2011) in the Briggs group subtomogram averaging pipeline and determined the first structures of incomplete Gag hexamers at the edges of the immature lattice. The pattern of conserved interfaces in these structures suggests that the basic unit of immature Gag lattice assembly is a dimer, and that partial CA-SP1 helical bundles are stable and may be important in maturation initiation.

In Chapter 4, I showed that a minimal system of mature, conical HIV-1 CA assemblies with the pocket factor IP₆ can be used to efficiently determine high-resolution CA hexamer and viral CA pentamer structures. I then used these assemblies to determine structures of the hexamer and pentamer bound to peptides derived from the nuclear import cofactors Nup153 and CPSF6, which I showed to bind to the hexamer but not the pentamer, in which part of the binding pocket is exposed due to a rotation between CA protomers.

5.1 Structural switch controlling HIV-1 maturation

Previous work in our group by de Marco et al. (2010) resulted in low-resolution structures of immature Gag hexamers in various Gag cleavage mutants, in which different combinations of Gag proteolytic cleavage sites were inactivated by mutation. In this work, I collected much larger data sets than de Marco et al. (2010) which were of much better quality due to the use of a counting direct electron detector (Li et al., 2013), dose-symmetric tilt scheme (Hagen et al., 2017) and improved data processing techniques such as dose-dependent amplitude weighting (Grant & Grigorieff, 2015) and 3D CTF correction (Turoňová et al., 2017). This enabled me to detect minority phenotypes which show that, contrary to previous evidence, structural maturation can occur when only one end of the CA-SP1 module in Gag undergoes proteolytic cleavage, albeit with drastically reduced efficiency. Notably, no single proteolytic cleavage was absolutely necessary for structural maturation, indicating that the relationship between proteolytic and structural maturation is complex.

I then determined structures of the CA hexamer in both the immature and mature lattices observed in the different cleavage mutant constructs to high resolution, which enabled a structural analysis of how each proteolytic cleavage affects structural maturation. Notably, when proteolytic cleavage was allowed to occur between MA and CA but not between CA, SP1 and NC, the structure of the CA hexamer in the immature lattice contained an extra density that corresponds to the beta-hairpin normally only found within the mature lattice. Conversely, in some viruses in which cleavage between MA and CA was blocked, structural maturation was able to proceed and the mature CA hexamer structure from these viruses was clearly lacking a beta-hairpin density. This indicates that the N-terminal beta-hairpin is structurally compatible with the immature lattice and dispensable for structural maturation, and thus cannot be the main determinant of this process. In contrast, presence of the CA-SP1 helical bundle was strictly correlated with immature lattice phenotype regardless of proteolytic cleavage state, and the corresponding residues were always disordered even when the cleavage site between CA and SP1 was mutated.

Taken together, these results support a model for maturation in which immature lattice stability is maintained by cumulative MA-membrane and NC-RNA interactions, with proteolytic cleavage at either end of CA sufficient to dissolve enough of these interactions for some structural maturation to occur, but with efficient maturation requiring cleavage at both ends of CA. The beta-hairpin is not sufficient to destabilise the immature lattice and trigger maturation, and it is instead dissolution of the CA-SP1 helical bundle that represents the main structural switch for maturation, with cleavage between CA and SP1 making the transition permanent.

5.2 HIV-1 lattice assembly and maturation initiation

HIV-1 maturation and acquisition of infectivity are sensitive to defects in viral assembly and budding (Bendjennat & Saffarian, 2016; Carlson et al., 2010), and an understanding of the assembly process can therefore be informative in studying the initiation of structural maturation. Here, I have presented the results of subtomogram classification of Gag hexamers at the edges of immature lattices from a previously published data set of wild-type, immature HIV-1 viruses (Schur et al., 2016). The existing classification implementation in the Briggs group subtomogram pipeline when I started this work, which was the same as that in the Dynamo software package (Castaño-Díez et al., 2012), was not effective at classifying this data set due to missing wedge effects. I therefore implemented an improved approach in our subtomogram averaging software in the form of the WMD PCA algorithm described by Heumann et al. (2011), and combined this with an improved method to calculate Fourier weights simultaneously developed in our group.

I then showed that this improved classification workflow was able to successfully classify hexamers at the edge of the Gag lattice based on absence or presence of neighbouring hexamers, which revealed the existence of incomplete hexamers at these lattice edges. The structures I obtained represent the first direct evidence that the immature HIV-1 Gag hexamer can exist in stable, incomplete conformations lacking between one and three Gag subunits from the hexamer. The pattern of interfaces remaining in these structures shows that Gag molecules are absent from the incomplete hexamers as dimers, and is inconsistent with a trimer being the basic multimeric state in which Gag is added to the lattice, therefore suggesting that Gag dimers are the basic unit of immature lattice assembly.

Another surprising result was the presence of ordered density in the CA-SP1 junction region consistent with a helical bundle structure, even when up to three Gag molecules were absent from the hexamer, which showed that partial helical bundles of CA-SP1 can be ordered. The resolution of the maps I obtained was insufficient to unambiguously assign positions of the remaining CA-SP1 helices, and therefore I collaborated with Alexander Pak (University of Chicago, Illinois, United States of America) in order to run molecular dynamics simulations of partial helical bundles. Consistent with existing evidence from NMR spectroscopy (Wang et al., 2017), interpreting my cryo-EM structures here in the context of these simulations suggests that partial helical bundles exist in a helix-coil equilibrium as seen with the complete six-helix bundle, but that partial bundles are considerably more prone to destabilisation. This suggests a role for partial helical bundles in incomplete hexamers as potential sites of maturation initiation, which could begin at the lattice edges and lead to progressive destabilisation of the immature lattice as part of structural maturation.

5.3 Interaction of the capsid with host cofactors

Interaction of the mature HIV-1 capsid with host cofactors is central to its roles in protecting the viral genome from degradation, nuclear import and integration of the provirus into host chromatin. The cofactor IP₆ has become prominent in recent years as it has been shown to act as a pocket factor that greatly enhances the stability of the capsid to uncoating (Mallery et al., 2018). Here, I used conical, capsid-like cores of recombinant HIV-1 CA protein assembled with IP₆ by my collaborator Robert Dick (Cornell University, Ithaca, New York, United States of America) to investigate binding of IP₆ and the nuclear import cofactors Nup153 and CPSF6 to both pentamers and hexamers across the curvature range of the capsid. Using subtomogram averaging, I demonstrated that the CA pentamer incorporated into these VLPs is the viral pentamer as opposed to the completely different crystal form of the pentamer, and that this system reproduces the structural properties of native viral capsids in a way that is amenable to cofactor-binding experiments.

My high-resolution hexamer and pentamer structures also demonstrated that IP₆ binds to the viral pentamer in a similar manner to the hexamer, and in both cases this is consistent with two IP₆ molecules in the N-terminal pore. I also extended the lattice curvature analysis approach using hexamer tilt and twist angle pairs, originally described by Mattei et al. (2016a), by using principal component analysis to describe the range of all unique local curvatures in terms of two eigenvectors. This enabled me to generate asymmetric structures of the hexamer with different unique curvatures, which verified that local lattice curvature does not appear to change the occupancy of the IP₆ binding pore in the hexamer.

I next determined high-resolution hexamer and pentamer structures with the peptides Nup153₁₄₀₇₋₁₄₂₃ and CPSF6₃₁₃₋₃₂₇. These show that the 19° rotation between CA protomers in the pentamer relative to the hexamer results in no observable peptide binding to the pentamer at a concentration of 50 µM, whereas the binding of both peptides to the hexamer is identical to that seen in published crystal structures (Price et al., 2014). I also applied hexamer tilt and twist angle pair analysis in order to investigate whether the binding occupancy of either of these peptides in the hexamer changes with lattice curvature. There was no significant change in CPSF6₃₁₃₋₃₂₇ occupancy with lattice curvature, but a marked reduction in occupancy of the Nup153₁₄₀₇₋₁₄₂₃ residues that bind to the neighbouring CA molecule at low tilt angles. This is consistent with the more cyclic structure of CPSF6₃₁₃₋₃₂₇ compared to the linear Nup153₁₄₀₇₋₁₄₂₃ that extends further out of the hydrophobic binding pocket shared by both peptides, and suggests that Nup153 occupancy is higher at highly-curved regions of the capsid than in flatter regions. This has potentially important implications for the exchange of these host cofactors during different parts of the viral life cycle, which may be modulated by subtle differences in occupancy.

5.4 Perspectives: immature lattice assembly, maturation initiation and capsid stability

Much progress has been made in recent years towards elucidating the molecular mechanisms of immature lattice assembly, initiation of capsid maturation and disassembly of the mature capsid within the host cell, and the results I have presented in this thesis raise further questions about these processes that need to be addressed. The work on immature Gag lattice edges that I have done in Chapter 3 provides direct evidence from incomplete hexamer structures that the dimer is most likely the basic unit of immature lattice assembly. This is consistent with computational models of lattice assembly (Pak et al., 2017), but it is still not clear how Gag-Pol incorporation into the lattice would occur in such a model. Proteolytic maturation requires dimerisation of the PR domain in Gag-Pol and activation of the protease via an autocatalytic cleavage step (Pettit et al., 2004), and given the 1:20 ratio of Gag-Pol:Gag incorporated into virus, successful maturation would require at least some of the Gag-Pol molecules to be incorporated close enough together during assembly for dimerisation to occur. This dimerisation would presumably need to occur either after budding or during the late stages of viral assembly, and thus within the context of an already-assembled Gag lattice, as premature proteolytic cleavage leads to unusually complete immature lattices and loss of the viral genome as seen in a subset of the viruses presented in Chapter 2. Attempts to locate Gag-Pol by cryo-ET in order to map its distribution within the immature virus have so far been unsuccessful, and further improvements in detector technology and data collection methodology will hopefully enable this problem to be approached in the future as it represents a key step in the initiation of maturation that is currently not fully understood.

The structural changes in the capsid that lead to uncoating within the host cell as well as the host cofactor interactions that modulate this process represent the next important questions that need to be fully addressed, as this stage of the viral life cycle is ripe for exploitation as a therapeutic target. Advances in cryo-ET data collection methodology and image processing, especially in classification approaches that more effectively deal with the problem of the missing wedge in tomographic data sets, mean that a structural approach to these problems is now feasible, as I have demonstrated in Chapter 3 for the immature lattice and Chapter 4 for the mature capsid. Extending the lattice edge classification approach I described in Chapter 3 to a subset of the mature CA cores from Chapter 4 which appear incomplete promises to provide important information about how disassembly of the capsid is achieved, which could be deduced from the pattern of interactions at broken lattice edges in a manner analogous to the approach I have described here for the immature lattice. Recently, Summers et al. (2019) published a method to engineer

HIV-1 CA protein *in vitro* into isolated, stable fragments that may emulate broken parts of the mature lattice. This represents another very promising, complementary approach to cryo-ET studies on mediation of capsid disassembly by host cofactors.

Related to studies of capsid stability, important questions also remain about recognition of the capsid by the nuclear pore, the mechanism by which these cofactors facilitate nuclear import and to what extent uncoating must occur for nuclear import. I have shown that, at the concentrations I tested, Nup153 and CPSF6 do not visibly interact with the pentamer in capsid-like VLPs. The question whether there exist other cofactors that bind to the pentamer remains to be more comprehensively addressed. The identification of a nuclear-associated cofactor bound to pentamers would have important implications for current models of capsid uncoating and nuclear recognition. There is also the question of what role competition of different cofactors that bind to this pocket plays in the viral life cycle. It is tempting to speculate that subtle changes in binding affinity based on the local curvature of the capsid result in differential binding of different cofactors to specific parts of the capsid, and that competition may facilitate a ‘hand off’ between different cofactors during the various stages of nuclear import. The VLP system which I used in Chapter 4 enables a high-throughput structural approach to this problem, as it is amenable to cofactor binding experiments and enables consistently high-resolution structural determination of viral pentamers with much less data than previously possible.

In situ cryo-ET approaches to studying the correlation between the extent of capsid uncoating and nuclear import also represent a key research direction that should be further explored, as this allows these processes to be dissected within the native environment of the viral capsid in a host cell. Correlative light and electron microscopy (CLEM) approaches to localise HIV-1 capsids within infected cells combined with focused ion beam (FIB) milling of these cells to enable *in situ* cryo-ET of capsids undergoing nuclear import are extremely promising and I predict that in the next few years, these techniques will greatly deepen our understanding of the nuclear import process and help to settle the ongoing debate in the field as to how much and where capsid disassembly occurs prior to nuclear import. The observation that most capsids in purified virus preparations are incomplete (Márquez et al., 2018; Mattei et al., 2016a) is also intriguing, and raises the question of how this is compatible with the protective function of the viral capsid within the host cell cytosol. The question of what fraction of HIV-1 capsids are actually complete shortly before and after fusion with a host cell still needs to be systematically explored in a more native system, and *in situ* cryo-ET approaches are well-suited to this task.

References

- Accola, M. A., Höglund, S., & Göttlinger, H. G. (1998). A putative α -helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *Journal of Virology*, 72(3), 2072–2078. Retrieved from <http://jvi.asm.org/content/72/3/2072>
- Accola, M. A., Strack, B., & Göttlinger, H. G. (2000). Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *Journal of Virology*, 74(12), 5395–5402. doi:10.1128/JVI.74.12.5395-5402.2000
- Achuthan, V., Ferreira, J. M., Sowd, G. A., Puray-Chavez, M., McDougall, W. M., Paulucci-Holthauzen, A., Wu, X., Fadel, H. J., Poeschla, E. M., Multani, A. S., Hughes, S. H., Sarafianos, S. G., Brass, A. L., & Engelman, A. N. (2018). Capsid-CPSF6 interaction licenses nuclear HIV-1 trafficking to sites of viral DNA integration. *Cell Host & Microbe*, 24(3), 392–404.e8. doi:10.1016/j.chom.2018.08.002
- Adamson, C. S., Ablan, S. D., Boeras, I., Goila-Gaur, R., Soheilian, F., Nagashima, K., Li, F., Salzwedel, K., Sakalian, M., Wild, C. T., & Freed, E. O. (2006). In vitro resistance to the human immunodeficiency virus type 1 maturation inhibitor PA-457 (bevirimat). *Journal of Virology*, 80(22), 10957–10971. doi:10.1128/JVI.01369-06
- Adrian, M., Dubochet, J., Lepault, J., & McDowell, A. W. (1984). Cryo-electron microscopy of viruses. *Nature*, 308(5954), 32–36. doi:10.1038/308032a0
- Ambrose, Z., Lee, K., Ndjomou, J., Xu, H., Oztop, I., Matous, J., Takemura, T., Unutmaz, D., Engelman, A., Hughes, S. H., & KewalRamani, V. N. (2012). Human immunodeficiency virus type 1 capsid mutation n74d alters cyclophilin a dependence and impairs macrophage infection. *Journal of Virology*, 86(8), 4708–4714. doi:10.1128/JVI.05887-11

- Anderson, E. M., & Maldarelli, F. (2018). The role of integration and clonal expansion in HIV infection: Live long and prosper. *Retrovirology*, *15*(1), 71. doi:10.1186/s12977-018-0448-8
- Aramburu, I. V., & Lemke, E. A. (2017). Floppy but not sloppy: Interaction mechanism of FG-nucleoporins and nuclear transport receptors. *Seminars in Cell & Developmental Biology*. Nuclear pores, *68*, 34–41. doi:10.1016/j.semcdb.2017.06.026
- Bai, X.-c., McMullan, G., & Scheres, S. H. W. (2015). How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, *40*(1), 49–57. doi:10.1016/j.tibs.2014.10.005
- Bartesaghi, A., Sprechmann, P., Liu, J., Randall, G., Sapiro, G., & Subramaniam, S. (2008). Classification and 3d averaging with missing wedge correction in biological electron tomography. *Journal of Structural Biology*, *162*(3), 436–450. doi:10.1016/j.jsb.2008.02.008
- Behrens, R. T., Aligeti, M., Pocock, G. M., Higgins, C. A., & Sherer, N. M. (2017). Nuclear export signal masking regulates HIV-1 rev trafficking and viral RNA nuclear export. *Journal of Virology*, *91*(3), e02107–16. doi:10.1128/JVI.02107-16
- Bejarano, D. A., Peng, K., Laketa, V., Börner, K., Jost, K. L., Lucic, B., Glass, B., Lusic, M., Müller, B., & Kräusslich, H.-G. (2019). HIV-1 nuclear import in macrophages is regulated by CPSF6-capsid interactions at the nuclear pore complex. *eLife*, *8*, e41800. doi:10.7554/eLife.41800
- Bendjennat, M., & Saffarian, S. (2016). The race against protease activation defines the role of ESCRTs in HIV budding. *PLOS Pathogens*, *12*(6), e1005657. doi:10.1371/journal.ppat.1005657
- Bharat, T. A. M., Davey, N. E., Ulbrich, P., Riches, J. D., de Marco, A., Rumlova, M., Sachse, C., Ruml, T., & Briggs, J. A. G. (2012). Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature*, *487*(7407), 385–389. doi:10.1038/nature11169
- Bhattacharya, A., Alam, S. L., Fricke, T., Zadrozny, K., Sedzicki, J., Taylor, A. B., Demeler, B., Pornillos, O., Ganser-Pornillos, B. K., Diaz-Griffero, F., Ivanov, D. N., & Yeager, M. (2014). Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proceedings of the National Academy of Sciences*, *111*(52), 18625–18630. doi:10.1073/pnas.1419945112
- Blair, W. S., Pickford, C., Irving, S. L., Brown, D. G., Anderson, M., Bazin, R., Cao, J., Ciaramella, G., Isaacson, J., Jackson, L., Hunt, R., Kjerrstrom, A., Nieman, J. A.,

- Patick, A. K., Perros, M., Scott, A. D., Whitby, K., Wu, H., & Butler, S. L. (2010). HIV capsid is a tractable target for small molecule therapeutic intervention. *PLOS Pathogens*, 6(12), e1001220. doi:10.1371/journal.ppat.1001220
- Blumenthal, R., Durell, S., & Viard, M. (2012). HIV entry and envelope glycoprotein-mediated fusion. *Journal of Biological Chemistry*, 287(49), 40841–40849. doi:10.1074/jbc.R112.406272
- Borland, L., & Heel, M. v. (1990). Classification of image data in conjugate representation spaces. *JOSA A*, 7(4), 601–610. doi:10.1364/JOSAA.7.000601
- Brenner, S., & Horne, R. W. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochimica et Biophysica Acta*, 34, 103–110. doi:10.1016/0006-3002(59)90237-9
- Briggs, J. A. G., & Kräusslich, H.-G. (2011). The molecular architecture of HIV. *Journal of Molecular Biology. Structural and Molecular Biology of HIV*, 410(4), 491–500. doi:10.1016/j.jmb.2011.04.021
- Briggs, J. A. G., Riches, J. D., Glass, B., Bartonova, V., Zanetti, G., & Kräusslich, H.-G. (2009). Structure and assembly of immature HIV. *Proceedings of the National Academy of Sciences*, 106(27), 11090–11095. doi:10.1073/pnas.0903535106
- Briggs, J. A. G., Wilk, T., Welker, R., Kräusslich, H.-G., & Fuller, S. D. (2003). Structural organization of authentic, mature HIV-1 virions and cores. *The EMBO Journal*, 22(7), 1707–1715. doi:10.1093/emboj/cdg143
- Buffone, C., Martinez-Lopez, A., Fricke, T., Opp, S., Severgnini, M., Cifola, I., Petiti, L., Frabetti, S., Skorupka, K., Zadrozny, K. K., Ganser-Pornillos, B. K., Pornillos, O., Nunzio, F. D., & Diaz-Griffero, F. (2018). Nup153 unlocks the nuclear pore complex for HIV-1 nuclear translocation in nondividing cells. *Journal of Virology*, 92(19), e00648–18. doi:10.1128/JVI.00648-18
- Burdick, R. C., Delviks-Frankenberry, K. A., Chen, J., Janaka, S. K., Sastri, J., Hu, W.-S., & Pathak, V. K. (2017). Dynamics and regulation of nuclear import and nuclear movements of HIV-1 complexes. *PLOS Pathogens*, 13(8), e1006570. doi:10.1371/journal.ppat.1006570
- Bykov, Y. S., Schaffer, M., Dodonova, S. O., Albert, S., Plitzko, J. M., Baumeister, W., Engel, B. D., & Briggs, J. A. (2017). The structure of the COPI coat determined within the cell. *eLife*, 6, e32493. doi:10.7554/eLife.32493

- Campbell, S., Fisher, R. J., Towler, E. M., Fox, S., Issaq, H. J., Wolfe, T., Phillips, L. R., & Rein, A. (2001). Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proceedings of the National Academy of Sciences*, 98(19), 10875–10879. doi:10.1073/pnas.191224698
- Carlson, L.-A., Briggs, J. A. G., Glass, B., Riches, J. D., Simon, M. N., Johnson, M. C., Müller, B., Grünewald, K., & Kräusslich, H.-G. (2008). Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host & Microbe*, 4(6), 592–599. doi:10.1016/j.chom.2008.10.013
- Carlson, L.-A., de Marco, A., Oberwinkler, H., Habermann, A., Briggs, J. A. G., Kräusslich, H.-G., & Grünewald, K. (2010). Cryo electron tomography of native HIV-1 budding sites. *PLoS Pathogens*, 6(11), e1001173. doi:10.1371/journal.ppat.1001173
- Caspar, D. L., & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposia on Quantitative Biology*, 27, 1–24. doi:10.1101/sqb.1962.027.001.005
- Castaño-Díez, D., Kudryashev, M., Arheit, M., & Stahlberg, H. (2012). Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *Journal of Structural Biology*. Special Issue: Electron Tomography, 178(2), 139–151. doi:10.1016/j.jsb.2011.12.017
- Checkley, M. A., Luttge, B. G., Soheilian, F., Nagashima, K., & Freed, E. O. (2010). The capsid-spacer peptide 1 Gag processing intermediate is a dominant-negative inhibitor of HIV-1 maturation. *Virology*, 400(1), 137–144. doi:10.1016/j.virol.2010.01.028
- Chen, B. (2019). Molecular mechanism of HIV-1 entry. *Trends in Microbiology*, 27(10), 878–891. doi:10.1016/j.tim.2019.06.002
- Coffin, J., & Swanstrom, R. (2013). HIV pathogenesis: Dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor Perspectives in Medicine*, 3(1), a012526. doi:10.1101/cshperspect.a012526
- Collier, D. A., Monit, C., & Gupta, R. K. (2019). The impact of HIV-1 drug escape on the global treatment landscape. *Cell Host & Microbe*, 26(1), 48–60. doi:10.1016/j.chom.2019.06.010
- Cosnefroy, O., Murray, P. J., & Bishop, K. N. (2016). HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology*, 13, 58. doi:10.1186/s12977-016-0292-7

- Crist, R. M., Datta, S. A. K., Stephen, A. G., Soheilian, F., Mirro, J., Fisher, R. J., Nagashima, K., & Rein, A. (2009). Assembly properties of human immunodeficiency virus type 1 Gag-leucine zipper chimeras: Implications for retrovirus assembly. *Journal of Virology*, 83(5), 2216–2225. doi:10.1128/JVI.02031-08
- Crowther, R. A., DeRosier, D. J., & Klug, A. (1970). The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 317(1530), 319–340. doi:10.1098/rspa.1970.0119
- Datta, S. A. K., Clark, P. K., Fan, L., Ma, B., Harvin, D. P., Sowder, R. C., Nussinov, R., Wang, Y.-X., & Rein, A. (2016). Dimerization of the SP1 region of HIV-1 Gag induces a helical conformation and association into helical bundles: Implications for particle assembly. *Journal of Virology*, 90(4), 1773–1787. doi:10.1128/JVI.02061-15
- Datta, S. A. K., Temeselew, L. G., Crist, R. M., Soheilian, F., Kamata, A., Mirro, J., Harvin, D., Nagashima, K., Cachau, R. E., & Rein, A. (2011). On the role of the SP1 domain in HIV-1 particle assembly: A molecular switch? *Journal of Virology*, 85(9), 4111–4121. doi:10.1128/JVI.00006-11
- de Marco, A., Heuser, A.-M., Glass, B., Kräusslich, H.-G., Müller, B., & Briggs, J. A. G. (2012). Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *Journal of Virology*, 86(24), 13708–13716. doi:10.1128/JVI.01704-12
- de Marco, A., Müller, B., Glass, B., Riches, J. D., Kräusslich, H.-G., & Briggs, J. A. G. (2010). Structural analysis of HIV-1 maturation using cryo-electron tomography. *PLoS Pathogens*, 6(11), e1001215. doi:10.1371/journal.ppat.1001215
- Delelis, O., Carayon, K., Saïb, A., Deprez, E., & Mouscadet, J.-F. (2008). Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology*, 5(1), 114. doi:10.1186/1742-4690-5-114
- DeRosier, D. J., & Klug, A. (1968). Reconstruction of three dimensional structures from electron micrographs. *Nature*, 217(5124), 130–134. doi:10.1038/217130a0
- Dettenhofer, M., & Yu, X.-F. (1999). Highly purified human immunodeficiency virus type 1 reveals a virtual absence of vif in virions. *Journal of Virology*, 73(2), 1460–1467. Retrieved from <http://jvi.asm.org/content/73/2/1460>
- Di Nunzio, F., Danckaert, A., Fricke, T., Perez, P., Fernandez, J., Perret, E., Roux, P., Shorte, S., Charneau, P., Diaz-Griffero, F., & Arhel, N. J. (2012). Human nucleo-

- porins promote HIV-1 docking at the nuclear pore, nuclear import and integration. *PLOS ONE*, 7(9), e46037. doi:10.1371/journal.pone.0046037
- Dick, R. A., Zadrozny, K. K., Xu, C., Schur, F. K. M., Lyddon, T. D., Ricana, C. L., Wagner, J. M., Perilla, J. R., Ganser-Pornillos, B. K., Johnson, M. C., Pornillos, O., & Vogt, V. M. (2018). Inositol phosphates are assembly co-factors for HIV-1. *Nature*, 1. doi:10.1038/s41586-018-0396-4
- Dodonova, S. O., Prinz, S., Bilanchone, V., Sandmeyer, S., & Briggs, J. A. G. (2019). Structure of the ty3/gypsy retrotransposon capsid and the evolution of retroviruses. *Proceedings of the National Academy of Sciences*, 201900931. doi:10.1073/pnas.1900931116
- Engel, B. D., Schaffer, M., Albert, S., Asano, S., Plitzko, J. M., & Baumeister, W. (2015). In situ structural analysis of golgi intracisternal protein arrays. *Proceedings of the National Academy of Sciences*, 112(36), 11264–11269. doi:10.1073/pnas.1515337112
- Erickson, H. P., & Klug, A. (1971). Measurement and compensation of defocusing and aberrations by fourier processing of electron micrographs. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 261(837), 105–118. doi:10.1098/rstb.1971.0040
- Erlendsson, S., Morado, D. R., Shepherd, J. D., & Briggs, J. A. G. (2019). Structures of virus-like capsids formed by the drosophila neuronal arc proteins. *bioRxiv*, 697193. doi:10.1101/697193
- Foley, B., Leitner, T., Apetrei, C., Hahn, B., Mizrachi, I., Mullins, J., Rambaut, A., Wolinsky, S., Korber, B., Singh, A., Abfalterer, W., Giori, E. E., Fischer, W., Hraber, P., Macke, J., Szinger, J. J., Wagh, K., & Yoon, H. (2018). HIV sequence compendium 2018, 448.
- Forshey, B. M., Schwedler, U. v., Sundquist, W. I., & Aiken, C. (2002). Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *Journal of Virology*, 76(11), 5667–5677. doi:10.1128/JVI.76.11.5667-5677.2002
- Förster, F., Medalia, O., Zauberman, N., Baumeister, W., & Fass, D. (2005). Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proceedings of the National Academy of Sciences of the United States of America*, 102(13), 4729–4734. doi:10.1073/pnas.0409178102

- Förster, F., Pruggnaller, S., Seybert, A., & Frangakis, A. S. (2008). Classification of cryo-electron sub-tomograms using constrained correlation. *Journal of Structural Biology*. The 4th International Conference on Electron Tomography, 161(3), 276–286. doi:10.1016/j.jsb.2007.07.006
- Francis, A. C., & Melikyan, G. B. (2018). Single HIV-1 imaging reveals progression of infection through CA-dependent steps of docking at the nuclear pore, uncoating, and nuclear transport. *Cell Host & Microbe*, 23(4), 536–548.e6. doi:10.1016/j.chom.2018.03.009
- Frank, G. A., Narayan, K., Bess Jr, J. W., Del Prete, G. Q., Wu, X., Moran, A., Hartnell, L. M., Earl, L. A., Lifson, J. D., & Subramaniam, S. (2015). Maturation of the HIV-1 core by a non-diffusional phase transition. *Nature Communications*, 6, 5854. doi:10.1038/ncomms6854
- Frank, J. (2006). *Three-dimensional electron microscopy of macromolecular assemblies: Visualization of biological molecules in their native state*. New York: Oxford University Press.
- Franke, E. K., Yuan, H. E. H., & Luban, J. (1994). Specific incorporation of cyclophilin a into HIV-1 virions. *Nature*, 372(6504), 359–362. doi:10.1038/372359a0
- Freed, E. O. (2015). HIV-1 assembly, release and maturation. *Nature Reviews Microbiology*, 13(8), 484–496. doi:10.1038/nrmicro3490
- Gheysen, D., Jacobs, E., de Foresta, F., Thiriart, C., Francotte, M., Thines, D., & De Wilde, M. (1989). Assembly and release of HIV-1 precursor pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell*, 59(1), 103–112. doi:10.1016/0092-8674(89)90873-8
- Ghosh, A. K., Osswald, H. L., & Prato, G. (2016). Recent progress in the development of HIV-1 protease inhibitors for the treatment of HIV/AIDS. *Journal of Medicinal Chemistry*, 59(11), 5172–5208. doi:10.1021/acs.jmedchem.5b01697
- Gitti, R. K., Lee, B. M., Walker, J., Summers, M. F., Yoo, S., & Sundquist, W. I. (1996). Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science*, 273(5272), 231–235. doi:10.1126/science.273.5272.231
- Glaeser, R. M. (2016). Chapter two - specimen behavior in the electron beam. In R. A. Crowther (Ed.), *Methods in enzymology* (Vol. 579, pp. 19–50). The Resolution Revolution: Recent Advances In cryoEM. Academic Press. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0076687916300283>

- Glaeser, R. M. (2013). Invited review article: Methods for imaging weak-phase objects in electron microscopy. *Review of Scientific Instruments*, 84(11), 111101. doi:10.1063/1.4830355
- González, M. E. (2015). Vpu protein: The viroporin encoded by HIV-1. *Viruses*, 7(8), 4352–4368. doi:10.3390/v7082824
- Göttlinger, H. G., Sodroski, J. G., & Haseltine, W. A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences*, 86(15), 5781–5785. doi:10.1073/pnas.86.15.5781
- Grant, T., & Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife*, 4, e06980. doi:10.7554/eLife.06980
- Graves, M. C., Lim, J. J., Heimer, E. P., & Kramer, R. A. (1988). An 11-kDa form of human immunodeficiency virus protease expressed in escherichia coli is sufficient for enzymatic activity. *Proceedings of the National Academy of Sciences*, 85(8), 2449–2453. doi:10.1073/pnas.85.8.2449
- Greene, W. C. (2007). A history of AIDS: Looking back to see ahead. *European Journal of Immunology*, 37, S94–S102. doi:10.1002/eji.200737441
- Gres, A. T., Kirby, K. A., KewalRamani, V. N., Tanner, J. J., Pornillos, O., & Sarafianos, S. G. (2015). X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science*, 349(6243), 99–103. doi:10.1126/science.aaa5936
- Grime, J. M. A., Dama, J. F., Ganser-Pornillos, B. K., Woodward, C. L., Jensen, G. J., Yeager, M., & Voth, G. A. (2016). Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nature Communications*, 7, 11568. doi:10.1038/ncomms11568
- Gross, I., Hohenberg, H., Huckhagel, C., & Kräusslich, H.-G. (1998). N-terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles. *Journal of Virology*, 72(6), 4798–4810. Retrieved from <http://jvi.asm.org/content/72/6/4798>
- Guzman, R. N. D., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 ψ -RNA recognition element. *Science*, 279(5349), 384–388. doi:10.1126/science.279.5349.384

- Hagan, M. F., Elrad, O. M., & Jack, R. L. (2011). Mechanisms of kinetic trapping in self-assembly and phase transformation. *The Journal of Chemical Physics*, 135(10), 104115. doi:10.1063/1.3635775
- Hagen, W. J. H., Wan, W., & Briggs, J. A. G. (2017). Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of Structural Biology. Electron Tomography*, 197(2), 191–198. doi:10.1016/j.jsb.2016.06.007
- Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H.-D., & Garten, W. (1992). Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gpl60. *Nature*, 360(6402), 358–361. doi:10.1038/360358a0
- Harauz, G., & Van Heel, M. (1985). Direct 3d reconstruction from projections with initially unknown angles. In H. Lemke, M. L. Rhodes, C. C. Jaffee, & R. Felix (Eds.), *Computer assisted radiology / computergestützte radiologie* (pp. 649–653). Springer Berlin Heidelberg.
- Heumann, J. M., Hoenger, A., & Mastronarde, D. N. (2011). Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *Journal of Structural Biology*, 175(3), 288–299. doi:10.1016/j.jsb.2011.05.011
- Hrabe, T., Chen, Y., Pfeffer, S., Kuhn Cuellar, L., Mangold, A.-V., & Förster, F. (2012). PyTom: A python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *Journal of Structural Biology. Special Issue: Electron Tomography*, 178(2), 177–188. doi:10.1016/j.jsb.2011.12.003
- Hu, W. S., & Temin, H. M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. *Proceedings of the National Academy of Sciences*, 87(4), 1556–1560. doi:10.1073/pnas.87.4.1556
- Hu, W.-S., & Hughes, S. H. (2012). HIV-1 reverse transcription. *Cold Spring Harbor Perspectives in Medicine*, 2(10), a006882. doi:10.1101/cshperspect.a006882
- Huang, P.-T., Summers, B. J., Xu, C., Perilla, J. R., Malikov, V., Naghavi, M. H., & Xiong, Y. (2019). FEZ1 is recruited to a conserved cofactor site on capsid to promote HIV-1 trafficking. *Cell Reports*, S2211124719309830. doi:10.1016/j.celrep.2019.07.079
- Ishima, R., Torchia, D. A., Lynch, S. M., Gronenborn, A. M., & Louis, J. M. (2003). Solution structure of the mature HIV-1 protease monomer INSIGHT INTO THE TERTIARY FOLD AND STABILITY OF a PRECURSOR. *Journal of Biological Chemistry*, 278(44), 43311–43319. doi:10.1074/jbc.M307549200

- Ivanchenko, S., Godinez, W. J., Lampe, M., Kräusslich, H.-G., Eils, R., Rohr, K., Bräuchle, C., Müller, B., & Lamb, D. C. (2009). Dynamics of HIV-1 assembly and release. *PLOS Pathog*, 5(11), e1000652. doi:10.1371/journal.ppat.1000652
- Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 331(6153), 280–283. doi:10.1038/331280a0
- Jacques, D. A., McEwan, W. A., Hilditch, L., Price, A. J., Towers, G. J., & James, L. C. (2016). HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature*, 536(7616), 349–353. doi:10.1038/nature19098
- Jensen, G. J., & Kornberg, R. D. (2000). Defocus-gradient corrected back-projection. *Ultramicroscopy*, 84(1), 57–64. doi:10.1016/S0304-3991(00)00005-X
- Joshi, A., Nagashima, K., & Freed, E. O. (2006). Mutation of dileucine-like motifs in the human immunodeficiency virus type 1 capsid disrupts virus assembly, Gag-Gag interactions, Gag-membrane binding, and virion maturation. *Journal of Virology*, 80(16), 7939–7951. doi:10.1128/JVI.00355-06
- Kaplan, A. H., Zack, J. A., Knigge, M., Paul, D. A., Kempf, D. J., Norbeck, D. W., & Swanstrom, R. (1993). Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. *Journal of Virology*, 67(7), 4050–4055. Retrieved from <https://jvi.asm.org/content/67/7/4050>
- Karn, J. (2011). The molecular biology of HIV latency: Breaking and restoring the tat-dependent transcriptional circuit. *Current Opinion in HIV and AIDS*, 6(1), 4. doi:10.1097/COH.0b013e328340ffbb
- Keller, P. W., Huang, R. K., England, M. R., Waki, K., Cheng, N., Heymann, J. B., Craven, R. C., Freed, E. O., & Steven, A. C. (2013). A two-pronged structural analysis of retroviral maturation indicates that core formation proceeds by a disassembly-reassembly pathway rather than a displacive transition. *Journal of Virology*, 87(24), 13655–13664. doi:10.1128/JVI.01408-13
- Kogan, M., & Rappaport, J. (2011). HIV-1 accessory protein vpr: Relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology*, 8(1), 25. doi:10.1186/1742-4690-8-25
- Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., & Sigal, I. S. (1988). Active human immunodeficiency virus protease

- is required for viral infectivity. *Proceedings of the National Academy of Sciences*, 85(13), 4686–4690. doi:10.1073/pnas.85.13.4686
- Konvalinka, J., Litterst, M. A., Welker, R., Kottler, H., Rippmann, F., Heuser, A. M., & Kräusslich, H. G. (1995). An active-site mutation in the human immunodeficiency virus type 1 proteinase (PR) causes reduced PR activity and loss of PR-mediated cytotoxicity without apparent effect on virus maturation and infectivity. *Journal of Virology*, 69(11), 7180–7186. Retrieved from <http://jvi.asm.org/content/69/11/7180>
- Kräusslich, H. G., Fäcke, M., Heuser, A. M., Konvalinka, J., & Zentgraf, H. (1995). The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *Journal of Virology*, 69(6), 3407–3419. Retrieved from <http://jvi.asm.org/content/69/6/3407>
- Kremer, J. R., Mastronarde, D. N., & McIntosh, J. R. (1996). Computer visualization of three-dimensional image data using IMOD. *Journal of Structural Biology*, 116(1), 71–76. doi:10.1006/jsbi.1996.0013
- Kühlbrandt, W. (2014). The resolution revolution. *Science*, 343(6178), 1443–1444. doi:10.1126/science.1251652
- Kutluay, S. B., & Bieniasz, P. D. (2010). Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLOS Pathogens*, 6(11), e1001200. doi:10.1371/journal.ppat.1001200
- Kutluay, S. B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., & Bieniasz, P. D. (2014). Global changes in the RNA binding specificity of HIV-1 Gag regulate virion genesis. *Cell*, 159(5), 1096–1109. doi:10.1016/j.cell.2014.09.057
- Lahaye, X., Satoh, T., Gentili, M., Cerboni, S., Conrad, C., Hurbain, I., El Marjou, A., Lacabartz, C., Lelièvre, J.-D., & Manel, N. (2013). The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. *Immunity*, 39(6), 1132–1142. doi:10.1016/j.immuni.2013.11.002
- Lamorte, L., Titolo, S., Lemke, C. T., Goudreau, N., Mercier, J.-F., Wardrop, E., Shah, V. B., Schwedler, U. K. v., Langelier, C., Banik, S. S. R., Aiken, C., Sundquist, W. I., & Mason, S. W. (2013). Discovery of novel small-molecule HIV-1 replication inhibitors that stabilize capsid complexes. *Antimicrobial Agents and Chemotherapy*, 57(10), 4622–4631. doi:10.1128/AAC.00985-13
- Lampe, M., Briggs, J. A. G., Endress, T., Glass, B., Riegelsberger, S., Kräusslich, H.-G., Lamb, D. C., Bräuchle, C., & Müller, B. (2007). Double-labelled HIV-1 particles for

- study of virus–cell interaction. *Virology*, 360(1), 92–104. doi:10.1016/j.virol.2006.10.005
- Lee, K., Ambrose, Z., Martin, T. D., Oztop, I., Mulky, A., Julias, J. G., Vandegraaff, N., Baumann, J. G., Wang, R., Yuen, W., Takemura, T., Shelton, K., Taniuchi, I., Li, Y., Sodroski, J., Littman, D. R., Coffin, J. M., Hughes, S. H., Unutmaz, D., Engelman, A., & KewalRamani, V. N. (2010). Flexible use of nuclear import pathways by HIV-1. *Cell Host & Microbe*, 7(3), 221–233. doi:10.1016/j.chom.2010.02.007
- Lee, S.-K., Harris, J., & Swanstrom, R. (2009). A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an achilles heel in the virus life cycle. *Journal of Virology*, 83(17), 8536–8543. doi:10.1128/JVI.00317-09
- Letcher, A. J., Schell, M. J., & Irvine, R. F. (2008). Do mammals make all their own inositol hexakisphosphate? *Biochemical Journal*, 416(2), 263–270. doi:10.1042/BJ20081417
- Lever, A., Gottlinger, H., Haseltine, W., & Sodroski, J. (1989). Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *Journal of Virology*, 63(9), 4085–4087. Retrieved from <https://jvi.asm.org/content/63/9/4085>
- Li, S., Hill, C. P., Sundquist, W. I., & Finch, J. T. (2000). Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature*, 407(6802), 409–413. doi:10.1038/35030177
- Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., & Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, 10(6), 584–590. doi:10.1038/nmeth.2472
- Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A., & Venkatesan, S. (1986). Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. *Journal of Virology*, 60(2), 771–775. Retrieved from <https://jvi.asm.org/content/60/2/771>
- Lippincott-Schwartz, J., Freed, E. O., & van Engelenburg, S. B. (2017). A consensus view of ESCRT-mediated human immunodeficiency virus type 1 abscission. *Annual Review of Virology*, 4(0), 309–325.
- Liu, C., Perilla, J. R., Ning, J., Lu, M., Hou, G., Ramalho, R., Himes, B. A., Zhao, G., Bedwell, G. J., Byeon, I.-J., Ahn, J., Gronenborn, A. M., Prevelige, P. E., Rousso, I., Aiken, C., Polenova, T., Schulten, K., & Zhang, P. (2016). Cyclophilin a stabilizes

- the HIV-1 capsid through a novel non-canonical binding site. *Nature Communications*, 7, 10714. doi:10.1038/ncomms10714
- Lu, K., Heng, X., & Summers, M. F. (2011). Structural determinants and mechanism of HIV-1 genome packaging. *Journal of Molecular Biology. Structural and Molecular Biology of HIV*, 410(4), 609–633. doi:10.1016/j.jmb.2011.04.029
- Lučić, V., Förster, F., & Baumeister, W. (2005). STRUCTURAL STUDIES BY ELECTRON TOMOGRAPHY: From cells to molecules. *Annual Review of Biochemistry*, 74(1), 833–865. doi:10.1146/annurev.biochem.73.011303.074112
- Lundquist, C. A., Tobiume, M., Zhou, J., Unutmaz, D., & Aiken, C. (2002). Nef-mediated downregulation of CD4 enhances human immunodeficiency virus type 1 replication in primary t lymphocytes. *Journal of Virology*, 76(9), 4625–4633. doi:10.1128/JVI.76.9.4625-4633.2002
- Malikov, V., da Silva, E. S., Jovasevic, V., Bennett, G., de Souza Aranha Vieira, D. A., Schulte, B., Diaz-Griffero, F., Walsh, D., & Naghavi, M. H. (2015). HIV-1 capsids bind and exploit the kinesin-1 adaptor FEZ1 for inward movement to the nucleus. *Nature Communications*, 6, 6660. doi:10.1038/ncomms7660
- Mallery, D. L., Márquez, C. L., McEwan, W. A., Dickson, C. F., Jacques, D. A., Anandapadamanaban, M., Bichel, K., Towers, G. J., Saiardi, A., Böcking, T., & James, L. C. (2018). IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife*, 7, e35335. doi:10.7554/eLife.35335
- Márquez, C. (2019). *Differential modulation of HIV-1 capsid uncoating kinetics revealed by a novel single-particle uncoating assay* (Doctoral dissertation, University of New South Wales, Sydney, Australia). Retrieved from https://www.unsw.edu.au/permalink/f/a5fmj0/unsworks_57857
- Márquez, C., Lau, D., Walsh, J., Shah, V., McGuinness, C., Wong, A., Aggarwal, A., Parker, M. W., Jacques, D. A., Turville, S., & Böcking, T. (2018). Kinetics of HIV-1 capsid uncoating revealed by single-molecule analysis. *eLife*, 7. doi:10.7554/eLife.34772
- Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*, 152(1), 36–51. doi:10.1016/j.jsb.2005.07.007
- Mattei, S. (2016). *Structural studies of HIV-1 maturation by cryo-electron tomography* (Doctoral dissertation, Ruperto-Carola University of Heidelberg).

- Mattei, S., Flemming, A., Anders-Össwein, M., Kräusslich, H.-G., Briggs, J. A. G., & Müller, B. (2015). RNA and nucleocapsid are dispensable for mature HIV-1 capsid assembly. *Journal of Virology*, 89(19), 9739–9747. doi:10.1128/JVI.00750-15
- Mattei, S., Glass, B., Hagen, W. J. H., Kräusslich, H.-G., & Briggs, J. A. G. (2016a). The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science*, 354(6318), 1434–1437. doi:10.1126/science.aah4972
- Mattei, S., Schur, F. K., & Briggs, J. A. (2016b). Retrovirus maturation — an extraordinary structural transformation. *Current Opinion in Virology*. Antiviral strategies • Virus structure and expression, 18, 27–35. doi:10.1016/j.coviro.2016.02.008
- Mattei, S., Tan, A., Glass, B., Müller, B., Kräusslich, H.-G., & Briggs, J. A. G. (2018). High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proceedings of the National Academy of Sciences*, 201811237. doi:10.1073/pnas.1811237115
- McEwen, B. F., & Marko, M. (2001). The emergence of electron tomography as an important tool for investigating cellular ultrastructure. *Journal of Histochemistry & Cytochemistry*, 49(5), 553–563. doi:10.1177/002215540104900502
- Meents, A., Gutmann, S., Wagner, A., & Schulze-Bries, C. (2010). Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures. *Proceedings of the National Academy of Sciences*, 107(3), 1094–1099. doi:10.1073/pnas.0905481107
- Müller, B., Anders, M., Akiyama, H., Welsch, S., Glass, B., Nikovics, K., Clavel, F., Tervo, H.-M., Keppler, O. T., & Kräusslich, H.-G. (2009). HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. *Journal of Biological Chemistry*, 284(43), 29692–29703. doi:10.1074/jbc.M109.027144
- Nakai, M., & Goto, T. (1996). Ultrastructure and morphogenesis of human immunodeficiency virus. *Journal of Electron Microscopy*, 45(4), 247–257. doi:10.1093/oxfordjournals.jmicro.a023441
- Nickell, S., Förster, F., Linaroudis, A., Net, W. D., Beck, F., Hegerl, R., Baumeister, W., & Plitzko, J. M. (2005). TOM software toolbox: Acquisition and analysis for electron tomography. *Journal of Structural Biology*, 149(3), 227–234. doi:10.1016/j.jsb.2004.10.006
- Ning, J., Erdemci-Tandogan, G., Yufenyuy, E. L., Wagner, J., Himes, B. A., Zhao, G., Aiken, C., Zandi, R., & Zhang, P. (2016). In vitro protease cleavage and computer

- simulations reveal the HIV-1 capsid maturation pathway. *Nature Communications*, 7, 13689. doi:10.1038/ncomms13689
- Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F., & Hance, A. J. (2007). Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. *Journal of Virology*, 81(14), 7620–7628. doi:10.1128/JVI.00083-07
- Novikova, M., Zhang, Y., Freed, E. O., & Peng, K. (2019). Multiple roles of HIV-1 capsid during the virus replication cycle. *Virologica Sinica*, 34(2), 119–134. doi:10.1007/s12250-019-00095-3
- Obr, M., & Schur, F. K. M. (2019). Chapter five - structural analysis of pleomorphic and asymmetric viruses using cryo-electron tomography and subtomogram averaging. In *Advances in virus research* (Vol. 105, pp. 117–159). doi:10.1016/bs.aivir.2019.07.008
- Ono, A., Ablan, S. D., Lockett, S. J., Nagashima, K., & Freed, E. O. (2004). Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proceedings of the National Academy of Sciences*, 101(41), 14889–14894. doi:10.1073/pnas.0405596101
- Ono, A., Waheed, A. A., Joshi, A., & Freed, E. O. (2005). Association of human immunodeficiency virus type 1 Gag with membrane does not require highly basic sequences in the nucleocapsid: Use of a novel Gag multimerization assay. *Journal of Virology*, 79(22), 14131–14140. doi:10.1128/JVI.79.22.14131-14140.2005
- Pak, A. J., Grime, J. M. A., Sengupta, P., Chen, A. K., Durumeric, A. E. P., Srivastava, A., Yeager, M., Briggs, J. A. G., Lippincott-Schwartz, J., & Voth, G. A. (2017). Immature HIV-1 lattice assembly dynamics are regulated by scaffolding from nucleic acid and the plasma membrane. *Proceedings of the National Academy of Sciences*, 201706600. doi:10.1073/pnas.1706600114
- Passmore, L. A., & Russo, C. J. (2016). Chapter three - specimen preparation for high-resolution cryo-EM. In R. A. Crowther (Ed.), *Methods in enzymology* (Vol. 579, pp. 51–86). The Resolution Revolution: Recent Advances In cryoEM. Academic Press. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0076687916300295>
- Passos, D. O., Li, M., Yang, R., Rebensburg, S. V., Ghirlando, R., Jeon, Y., Shkriabai, N., Kvaratskhelia, M., Craigie, R., & Lyumkis, D. (2017). Cryo-EM structures and

- atomic model of the HIV-1 strand transfer complex intasome. *Science*, 355(6320), 89–92. doi:10.1126/science.aah5163
- Peet, M. J., Henderson, R., & Russo, C. J. (2019). The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules. *Ultramicroscopy*. 75th Birthday of Christian Colliex, 85th Birthday of Archie Howie, and 75th Birthday of Hannes Lichte / PICO 2019 - Fifth Conference on Frontiers of Aberration Corrected Electron Microscopy, 203, 125–131. doi:10.1016/j.ultramic.2019.02.007
- Permanyer, M., Ballana, E., & Esté, J. A. (2010). Endocytosis of HIV: Anything goes. *Trends in Microbiology*, 18(12), 543–551. doi:10.1016/j.tim.2010.09.003
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. doi:10.1002/jcc.20084
- Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M., & Kaplan, A. H. (2004). Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *Journal of Virology*, 78(16), 8477–8485. doi:10.1128/JVI.78.16.8477-8485.2004
- Pettit, S. C., Lindquist, J. N., Kaplan, A. H., & Swanstrom, R. (2005). Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-pro-pol precursor are cleaved by the viral protease at different rates. *Retrovirology*, 2, 66. doi:10.1186/1742-4690-2-66
- Pornillos, O., & Ganser-Pornillos, B. K. (2019). Maturation of retroviruses. *Current Opinion in Virology*. Virus structure and expression, 36, 47–55. doi:10.1016/j.coviro.2019.05.004
- Pornillos, O., Ganser-Pornillos, B. K., & Yeager, M. (2011). Atomic-level modelling of the HIV capsid. *Nature*, 469(7330), 424–427. doi:10.1038/nature09640
- Price, A. J., Fletcher, A. J., Schaller, T., Elliott, T., Lee, K., KewalRamani, V. N., Chin, J. W., Towers, G. J., & James, L. C. (2012). CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLOS Pathogens*, 8(8), e1002896. doi:10.1371/journal.ppat.1002896
- Price, A. J., Jacques, D. A., McEwan, W. A., Fletcher, A. J., Essig, S., Chin, J. W., Halambage, U. D., Aiken, C., & James, L. C. (2014). Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. *PLOS Pathogens*, 10(10), e1004459. doi:10.1371/journal.ppat.1004459

- Qu, K., Glass, B., Doležal, M., Schur, F. K. M., Murciano, B., Rein, A., Rumlová, M., Ruml, T., Kräusslich, H.-G., & Briggs, J. A. G. (2018). Structure and architecture of immature and mature murine leukemia virus capsids. *Proceedings of the National Academy of Sciences*, 201811580. doi:10.1073/pnas.1811580115
- Radermacher, M. (1992). Weighted back-projection methods. In J. Frank (Ed.), *Electron tomography: Three-dimensional imaging with the transmission electron microscope* (pp. 91–115). doi:10.1007/978-1-4757-2163-8_5
- Ramalho, R., Rankovic, S., Zhou, J., Aiken, C., & Rousso, I. (2016). Analysis of the mechanical properties of wild type and hyperstable mutants of the HIV-1 capsid. *Retrovirology*, 13(1), 17. doi:10.1186/s12977-016-0250-4
- Rankovic, S., Ramalho, R., Aiken, C., & Rousso, I. (2018). PF74 reinforces the HIV-1 capsid to impair reverse transcription-induced uncoating. *Journal of Virology*, JVI.00845–18. doi:10.1128/JVI.00845-18
- Rankovic, S., Varadarajan, J., Ramalho, R., Aiken, C., & Rousso, I. (2017). Reverse transcription mechanically initiates HIV-1 capsid disassembly. *Journal of Virology*, JVI.00289–17. doi:10.1128/JVI.00289-17
- Rapaport, D. C. (2008). Role of reversibility in viral capsid growth: A paradigm for self-assembly. *Physical Review Letters*, 101(18), 186101. doi:10.1103/PhysRevLett.101.186101
- Rasaiyaah, J., Tan, C. P., Fletcher, A. J., Price, A. J., Blondeau, C., Hilditch, L., Jacques, D. A., Selwood, D. L., James, L. C., Noursadeghi, M., & Towers, G. J. (2013). HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature*, 503(7476), 402–405. doi:10.1038/nature12769
- Reddy, V. S., Natchiar, S. K., Gritton, L., Mullen, T. .-M., Stewart, P. L., & Nemerow, G. R. (2010a). Crystallization and preliminary x-ray diffraction analysis of human adenovirus. *Virology*, 402(1), 209–214. doi:10.1016/j.virol.2010.03.028
- Reddy, V. S., Natchiar, S. K., Stewart, P. L., & Nemerow, G. R. (2010b). Crystal structure of human adenovirus at 3.5 Å resolution. *Science*, 329(5995), 1071–1075. doi:10.1126/science.1187292
- Rice, W. J., Cheng, A., Noble, A. J., Eng, E. T., Kim, L. Y., Carragher, B., & Potter, C. S. (2018). Routine determination of ice thickness for cryo-EM grids. *Journal of Structural Biology*, 204(1), 38–44. doi:10.1016/j.jsb.2018.06.007

- Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*. Recent Advances in Detector Technologies and Applications for Molecular TEM, *192*(2), 216–221. doi:10.1016/j.jsb.2015.08.008
- Rosé, S., Hensley, P., O'Shannessy, D. J., Culp, J., Debouck, C., & Chaiken, I. (1992). Characterization of HIV-1 p24 self-association using analytical affinity chromatography. *Proteins: Structure, Function, and Bioinformatics*, *13*(2), 112–119. doi:10.1002/prot.340130204
- Saad, J. S., Miller, J., Tai, J., Kim, A., Ghanam, R. H., & Summers, M. F. (2006). Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proceedings of the National Academy of Sciences*, *103*(30), 11364–11369. doi:10.1073/pnas.0602818103
- Scheres, S. H. W. (2012). RELION: Implementation of a bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530. doi:10.1016/j.jsb.2012.09.006
- Scheres, S. H. W., Gao, H., Valle, M., Herman, G. T., Eggermont, P. P. B., Frank, J., & Carazo, J.-M. (2007). Disentangling conformational states of macromolecules in 3d-EM through likelihood optimization. *Nature Methods*, *4*(1), 27–29. doi:10.1038/nmeth992
- Schur, F. K. M., Dick, R. A., Hagen, W. J. H., Vogt, V. M., & Briggs, J. A. G. (2015). The structure of immature virus-like rous sarcoma virus Gag particles reveals a structural role for the p10 domain in assembly. *Journal of Virology*, *89*(20), 10294–10302. doi:10.1128/JVI.01502-15
- Schur, F. K. M., Hagen, W. J. H., de Marco, A., & Briggs, J. A. G. (2013). Determination of protein structure at 8.5 Å resolution using cryo-electron tomography and sub-tomogram averaging. *Journal of Structural Biology*, *184*(3), 394–400. doi:10.1016/j.jsb.2013.10.015
- Schur, F. K. M., Obr, M., Hagen, W. J. H., Wan, W., Jakobi, A. J., Kirkpatrick, J. M., Sachse, C., Kräusslich, H.-G., & Briggs, J. A. G. (2016). An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science*, *353*(6298), 506–508. doi:10.1126/science.aaf9620
- Sharkey, M., Sharova, N., Mohammed, I., Huff, S. E., Kummetha, I. R., Singh, G., Rana, T. M., & Stevenson, M. (2019). HIV-1 escape from small-molecule antagonism of vif. *mBio*, *10*(1), e00144–19. doi:10.1128/mBio.00144-19

- Sharp, P. M., & Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, 1(1), a006841. doi:10.1101/cshperspect.a006841
- Shi, J., Zhou, J., Shah, V. B., Aiken, C., & Whitby, K. (2011). Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. *Journal of Virology*, 85(1), 542–549. doi:10.1128/JVI.01406-10
- Singh, K., Gallazzi, F., Hill, K. J., Burke, D. H., Lange, M. J., Quinn, T. P., Neogi, U., & Sönnnerborg, A. (2019). GS-CA compounds: First-in-class HIV-1 capsid inhibitors covering multiple grounds. *Frontiers in Microbiology*, 10. doi:10.3389/fmicb.2019.01227
- Sommer, C., Straehle, C., Köthe, U., & Hamprecht, F. A. (2011). Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE international symposium on biomedical imaging: From nano to macro* (pp. 230–233). 2011 IEEE international symposium on biomedical imaging: From nano to macro. doi:10.1109/ISBI.2011.5872394
- Sowd, G. A., Serrao, E., Wang, H., Wang, W., Fadel, H. J., Poeschla, E. M., & Engelman, A. N. (2016). A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proceedings of the National Academy of Sciences*, 113(8), E1054–E1063. doi:10.1073/pnas.1524213113
- Stephenson, J. D., Li, H., Kenyon, J. C., Symmons, M., Klenerman, D., & Lever, A. M. L. (2013). Three-dimensional RNA structure of the major HIV-1 packaging signal region. *Structure*, 21(6), 951–962. doi:10.1016/j.str.2013.04.008
- Summers, B. J., Digianantonio, K. M., Smaga, S. S., Huang, P.-T., Zhou, K., Gerber, E. E., Wang, W., & Xiong, Y. (2019). Modular HIV-1 capsid assemblies reveal diverse host-capsid recognition mechanisms. *Cell Host & Microbe*, 26(2), 203–216.e6. doi:10.1016/j.chom.2019.07.007
- Summers, M. F., South, T. L., Kim, B., & Hare, D. R. (1990). High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry*, 29(2), 329–340. doi:10.1021/bi00454a005
- Sundquist, W. I., & Kräusslich, H.-G. (2012). HIV-1 assembly, budding, and maturation. *Cold Spring Harbor Perspectives in Medicine*, 2(7), a006924. doi:10.1101/cshperspect.a006924
- Tang, C., Loeliger, E., Luncsford, P., Kinde, I., Beckett, D., & Summers, M. F. (2004). Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proceedings of the National Academy of Sciences*, 101(2), 517–522. doi:10.1073/pnas.0305665101

- Taylor, K. A., & Glaeser, R. M. (1976). Electron microscopy of frozen hydrated biological specimens. *Journal of Ultrastructure Research*, 55(3), 448–456. doi:10.1016/S0022-5320(76)80099-8
- Tedbury, P. R., & Freed, E. O. (2015). HIV-1 Gag: An emerging target for antiretroviral therapy. In *The future of HIV-1 therapeutics* (pp. 171–201). Current Topics in Microbiology and Immunology. doi:10.1007/82_2015_436
- The PyMOL molecular graphics system, version 1.8.6.0 schrödinger, LLC. New York: Schrödinger, LLC.
- Tse, W., Link, J., Mulato, A., Niedziela-Majka, A., Rowe, W., Somoza, J., Villasenor, A., Yant, S., Zhang, J., & Zheng, J. (2017). Discovery of novel potent HIV capsid inhibitors with long-acting potential. In *Conference on retroviruses and opportunistic infections* (pp. 13–16).
- Turoňová, B., Schur, F. K. M., Wan, W., & Briggs, J. A. G. (2017). Efficient 3d-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *Journal of Structural Biology*, 199(3), 187–195. doi:10.1016/j.jsb.2017.07.007
- van Heel, M., & Frank, J. (1981). Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, 6(1), 187–194. doi:10.1016/S0304-3991(81)80197-0
- van Heel, M., Portugal, R. V., & Schatz, M. (2016). Multivariate statistical analysis of large datasets: Single particle electron microscopy. *Open Journal of Statistics*, 06(4), 701. doi:10.4236/ojs.2016.64059
- Villa, E., Schaffer, M., Plitzko, J. M., & Baumeister, W. (2013). Opening windows into the cell: Focused-ion-beam milling for cryo-electron tomography. *Current Opinion in Structural Biology*. Protein-carbohydrate interactions / Biophysical methods, 23(5), 771–777. doi:10.1016/j.sbi.2013.08.006
- von Schwedler, U. K., Stemmler, T. L., Klishko, V. Y., Li, S., Albertine, K. H., Davis, D. R., & Sundquist, W. I. (1998). Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *The EMBO Journal*, 17(6), 1555–1568. doi:10.1093/emboj/17.6.1555
- Voortman, L. M., Franken, E. M., van Vliet, L. J., & Rieger, B. (2012). Fast, spatially varying CTF correction in TEM. *Ultramicroscopy*, 118, 26–34. doi:10.1016/j.ultramic.2012.05.002

- Wade, R. H. (1992). A brief look at imaging and contrast transfer. *Ultramicroscopy*, 46(1), 145–156. doi:10.1016/0304-3991(92)90011-8
- Wagner, J. M., Zadrozny, K. K., Chrustowicz, J., Purdy, M. D., Yeager, M., Ganser-Pornillos, B. K., & Pornillos, O. (2016). Crystal structure of an HIV assembly and maturation switch. *eLife*, 5, e17063. doi:10.7554/eLife.17063
- Waki, K., Durell, S. R., Soheilian, F., Nagashima, K., Butler, S. L., & Freed, E. O. (2012). Structural and functional insights into the HIV-1 maturation inhibitor binding pocket. *PLOS Pathogens*, 8(11), e1002997. doi:10.1371/journal.ppat.1002997
- Wan, W., & Briggs, J. A. G. (2016). Chapter thirteen - cryo-electron tomography and subtomogram averaging. In R. A. Crowther (Ed.), *Methods in enzymology* (Vol. 579, pp. 329–367). The Resolution Revolution: Recent Advances In cryoEM. Academic Press. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0076687916300325>
- Wang, M., Quinn, C. M., Perilla, J. R., Zhang, H., Jr, R. S., Hou, G., Byeon, I.-J., Suiter, C. L., Ablan, S., Urano, E., Nitz, T. J., Aiken, C., Freed, E. O., Zhang, P., Schulten, K., Gronenborn, A. M., & Polenova, T. (2017). Quenching protein dynamics interferes with HIV capsid maturation. *Nature Communications*, 8(1), 1779. doi:10.1038/s41467-017-01856-y
- Wieggers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H., & Kräusslich, H.-G. (1998). Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *Journal of Virology*, 72(4), 2846–2854. Retrieved from <http://jvi.asm.org/content/72/4/2846>
- Willey, R. L., Bonifacino, J. S., Potts, B. J., Martin, M. A., & Klausner, R. D. (1988). Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. *Proceedings of the National Academy of Sciences*, 85(24), 9580–9584. doi:10.1073/pnas.85.24.9580
- Wu, Y. (2004). HIV-1 gene expression: Lessons from provirus and non-integrated DNA. *Retrovirology*, 1(1), 13. doi:10.1186/1742-4690-1-13
- Xiong, Q., Morphew, M. K., Schwartz, C. L., Hoenger, A. H., & Mastronarde, D. N. (2009). CTF determination and correction for low dose tomographic tilt series. *Journal of Structural Biology*, 168(3), 378–387. doi:10.1016/j.jsb.2009.08.016
- Yang, Y., Qu, N., Tan, J., Rushdi, M. N., Krueger, C. J., & Chen, A. K. (2018). Roles of Gag-RNA interactions in HIV-1 virus assembly deciphered by single-molecule

- localization microscopy. *Proceedings of the National Academy of Sciences*, 115(26), 6721–6726. doi:10.1073/pnas.1805728115
- Yu, X., Yuan, X., McLane, M. F., Lee, T. H., & Essex, M. (1993). Mutations in the cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein impair the incorporation of env proteins into mature virions. *Journal of Virology*, 67(1), 213–221. Retrieved from <https://jvi.asm.org/content/67/1/213>
- Yuan, X., Yu, X., Lee, T. H., & Essex, M. (1993). Mutations in the n-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. *Journal of Virology*, 67(11), 6387–6394. Retrieved from <https://jvi.asm.org/content/67/11/6387>
- Zhao, G., Perilla, J. R., Yufenyuy, E. L., Meng, X., Chen, B., Ning, J., Ahn, J., Groenborn, A. M., Schulten, K., Aiken, C., & Zhang, P. (2013). Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497(7451), 643–646. doi:10.1038/nature12162
- Zheng, J., Yant, S. R., Ahmadyar, S., Chan, T. Y., Chiu, A., Cihlar, T., Link, J. O., Lu, B., Mwangi, J., Rowe, W., Schroeder, S. D., Stepan, G. J., Wang, K. W., Subramanian, R., & Tse, W. C. (2018). 539. GS-CA2: A novel, potent, and selective first-in-class inhibitor of HIV-1 capsid function displays nonclinical pharmacokinetics supporting long-acting potential in humans. *Open Forum Infectious Diseases*, 5, S199–S200. doi:10.1093/ofid/ofy210.548
- Zhou, W., Parent, L. J., Wills, J. W., & Resh, M. D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *Journal of Virology*, 68(4), 2556–2569. Retrieved from <https://jvi.asm.org/content/68/4/2556>

Publications and Permissions

Chapter 2 includes text and figures originally produced by myself for publication in the following paper, on which I am co-first author:

***Mattei, S., *Tan, A.,** Glass, B., Müller, B., Kräusslich, H.-G., & Briggs, J. A. G. (2018). High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proceedings of the National Academy of Sciences*, 201811237. doi:10.1073/pnas.1811237115

*** Equal contribution**

The above work is Copyright © 2018 the Author(s) and Published by the Proceedings of the National Academy of Sciences as an open access article under a Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND). The relevant terms of this licence state, as of 2 July 2019:

“Section 2 – Scope.

1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:
 - (A) reproduce and Share the Licensed Material, in whole or in part, for NonCommercial purposes only; and
 - (B) produce and reproduce, but not Share, Adapted Material for NonCommercial purposes only.”

“Section 3 – License Conditions.

(a) Attribution.

- (1) If You Share the Licensed Material, You must:
 - (A) retain the following if it is supplied by the Licensor with the Licensed Material:
 - (i) identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);
 - (ii) a copyright notice;
 - (iii) a notice that refers to this Public License;
 - (iv) a notice that refers to the disclaimer of warranties;
 - (v) a URI or hyperlink to the Licensed Material to the extent reasonably practicable;
 - (B) indicate if You modified the Licensed Material and retain an indication of any previous modifications; and
 - (C) indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.

For the avoidance of doubt, You do not have permission under this Public License to Share Adapted Material.”

The PNAS Author Center (www.pnas.org/page/authors/licenses) states, as of 2 July 2019: “The author(s) retains copyright to individual PNAS articles, and the National Academy of Sciences of the United States of America (NAS) holds copyright to the collective work and retains an exclusive License to Publish these articles, except for open access articles submitted beginning September 2017.”

Acknowledgements

I would like to thank my PhD supervisor, Dr. John Briggs for giving me the opportunity to work with him and for all of the invaluable advice and guidance he has so patiently and generously given me during my time in his research group. It has been an honour to work in such a fantastic group, both at EMBL Heidelberg and the MRC LMB in Cambridge, and I am glad to have had the chance to work on so many exciting and challenging projects that have helped me to better myself as a scientist.

I would also like to thank the members of my Thesis Advisory Committee (TAC): Dr. Martin Beck, Prof. Dr. Britta Brügger and Dr. Gerard Kleywegt, for their constant support and advice throughout the course of my PhD. They have always been very supportive of me, especially during the move from Heidelberg to Cambridge and when I encountered difficulties during the first year of my PhD. Their input during my TAC meetings has also been extremely useful in planning how best to proceed in the various projects that I have done and for that I am very grateful. I also thank the additional members of my examination committee, Prof. Dr. Oliver Fackler and Dr. Anna Kreshuk.

I am especially grateful to all of my collaborators who have generously provided me with amazing samples that made the work in this thesis possible: Robert Dick, as well as members of the research groups of Prof. Dr. Hans-Georg Kräusslich and apl. Prof. Dr. Barbara Müller, especially Bärbel Glass and Martin Obr. I also thank my collaborator Alexander Pak for the molecular dynamics simulations he performed based on my immature Gag lattice edge structures.

I also want to thank all current and former members of the Briggs group with whom I have had the pleasure to work: Julia, Kun, Oleksiy, Dustin, Yura, Lauren Ann, Zunlong, Kat, Simon, Alex, Lucy, Kino, Kou, Wim, Florian, Simone, Will, Ori, Cecilia, Beata, Svetlana, Mairi, Andreas and Sumi. They have been extremely helpful and great fun to work with, and have made my time in the Briggs group very memorable. I am deeply grateful to Dustin for all the sleepless nights at the microscope and his help with cryo-EM data collection, as well as for the many discussions we have had about cryo-EM data

processing and image classification algorithms. I also thank Wim Hagen for all of his support with electron microscopy when we were based in EMBL Heidelberg.

I am also very grateful to the Briggs group members who moved with me from Heidelberg to Cambridge: Julia, Yura, Oleksiy, Kun and Lauren Ann, for the very friendly and supportive environment they created during the sometimes confusing initial setup period in Cambridge. I want to specially thank Zunlong and Julia for their critical reading of parts of my thesis, as well as Oleksiy, Simon and Alex for giving me so much advice on planning wet lab experiments. Kino, Dustin and Kun have also taught me a lot about the mathematics of image processing, for which I am very grateful. Zunlong, Julia, Alex and Kat have been great company during late nights in the lab and always have a seemingly endless supply of snacks to keep me motivated. Zunlong especially has brought a lot of craziness and fun to the group atmosphere and has always been ready to listen whenever I was going through difficult periods in my projects.

Most importantly, I thank my family both in Malaysia and in the UK, especially my parents, for always supporting me in my decisions and for being there whenever I needed them. Without their support, I would not have been able to achieve any of this.

